DEPLOYING AND SENSING THE AIR QUALITY INDEX USING INTERNET OF THINGS AND SPARK MODEL

T Rajpradeesh[1], Bathrinath S[2], Vairavel Madeshwaren[3], Dhanasekar M[4], A.Askar Hussain[5]

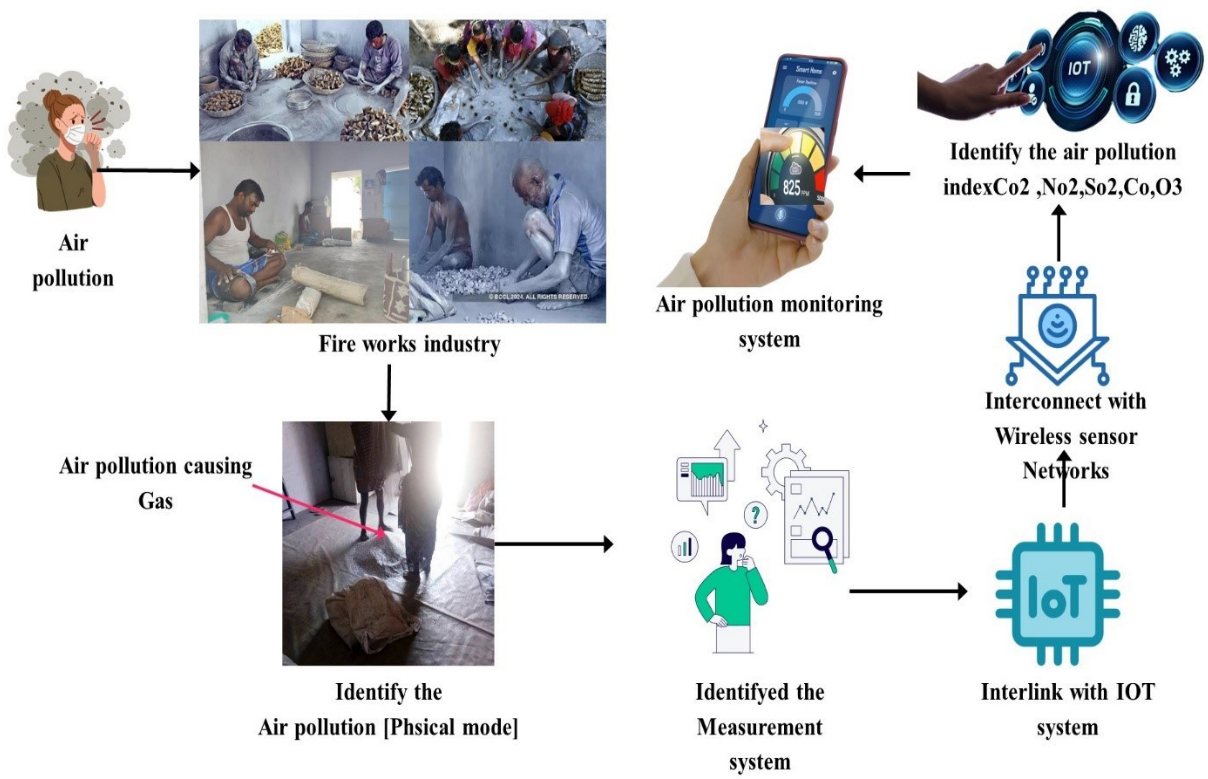
[*1,2,4] Department of Mechanical Engineering, Kalasalingam Academy of Research and Education, krishnankoil, Tamil Nadu, India.

[3]Professor, Department of Mechanical engineering, Annapoorna Engineering College, Salem, Tamil Nadu, India.

[5]Department of Industrial Safety Engineering, kalasalingam Academy of Research and Education, krishnankoil, Tamil Nadu, India.

*Corresponding mail: **pradeeshr132@gmail.com**

Graphical Abstract



Abstract

Air is crucial for human survival and a healthy existence, but in modern metropolitan living, it has become a dangerous issue due to its high pollution levels. When weighed against all other businesses, the fireworks sector is extremely dangerous. Every year, air pollution affects many people, including fireworkers, and causes various health problems that can occasionally result in death. Indeed, this work attempted to develop an accurate model for predicting air quality in the United States using a dataset obtained from linked Internet of Things (IoT) devices and Spark model, specifically wireless sensor networks (WSN). In order to predict air pollution caused by the introduction of hazardous substances including SO_2, NO_2, O_3,Particulate matter and CO into the Earth's atmosphere, this study explores the concept of merging the concepts of the Internet of Things. In conclusion, understanding forecast quality requires the computation of assessment measures using the proposed model. This research work used the RMSE to evaluate our predictions made using the Spark model. Good prediction models should typically have RMSE values of less than 0.3. It is accurate to say that our RMSE is $0.14 < 0.4$, which supports the validity of our model.

Keywords: Air pollution. Internet of Things, Fireworks, Wireless Sensor Networks, Spark model, RMSE value

1.Introduction

The release of contaminants into the atmosphere that are detrimental to both human health and the ecosystem at large is referred to as air pollution. Another way to describe it is as a shift in air quality that is assessed for biological, physical, or chemical pollutants. Therefore, the undesirable presence of pollutants or an unusual rise in the quantity of some atmospheric constituents is classified as air pollution. There are two categories of air pollution: invisible as well as visible. Compounds in the atmosphere are harmful to materials, the environment, and the

ACCEPTED MANUSCRIPT

health of people and animals. Air pollutants come in a variety of forms, such as gases, biological molecules, and both organic and inorganic particles. Consequently, the requirement for an automated method to forecast air pollution levels arises.

Our system's objective is to offer a mobile application that effectively provides the future air quality of a given place and notifies users in the case of severe air pollution. There are many different ingredients in fireworks. The overall quality of the air is lowered when these firecrackers burn because they release a significant amount of hazardous metals and gaseous and particle air pollutants. A study conducted in California, USA, discovered that following the Fourth of July celebration, there was a considerable increase in the ambient air's concentrations of magnesium, aluminium, potassium, lead, barium, strontium, and copper. The original chemical makeup and particle size of common firework combinations are covered in the paper. A thorough analysis of the air pollution caused by a fireworks display during Beijing's Lantern Day Festival revealed increases in SO_2, NO_2, and PM_{10} levels of 57,25, and 183% over the preceding day. It was found that the $PM_{2.5}$ concentration was six times more what it would be on an average day. Figure 1 shows the main components of air pollution.

ACCEPTED MANUSCRIPT

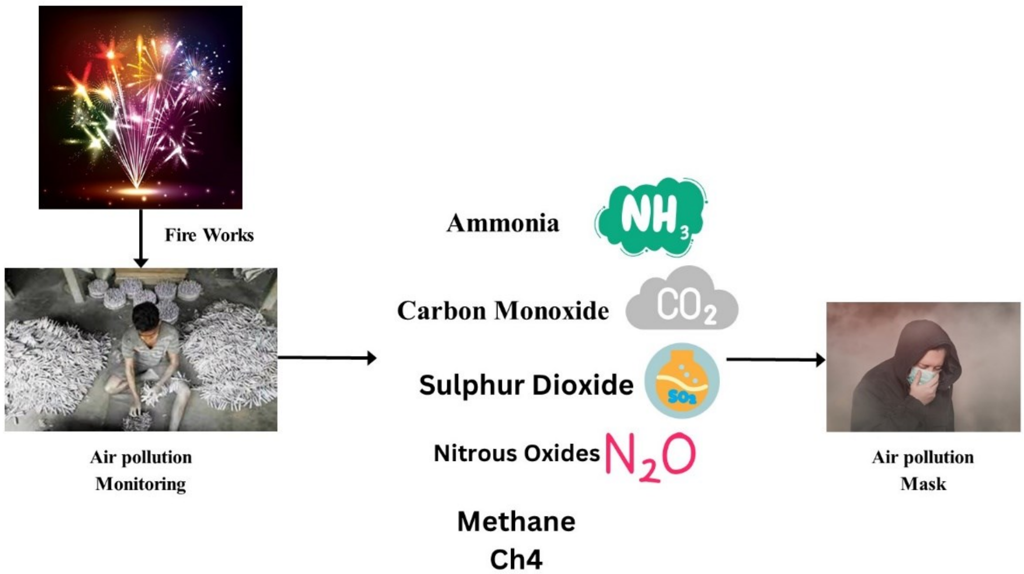


Figure 1: Components of air pollution

Air quality is a critical factor in human health and environmental sustainability, with high concentrations of pollutants such as particulate matter, nitrogen dioxide, sulfur dioxide, carbon monoxide, and ozone. The Internet of Things (IoT) has revolutionized air quality measurement by providing real-time data collection, analysis, and decision-making. IoT devices, such as low-cost sensors, drones, and mobile monitoring platforms, gather continuous and high-resolution data streams for comprehensive assessments of pollution levels. They can be deployed in various settings, including urban areas, industrial sites, transportation networks, and indoor spaces.

IoT technologies facilitate the integration of air quality data with other sources of information, enabling advanced analytics techniques to derive actionable insights and predictive models for air quality management. These data-driven approaches enable stakeholders to identify pollution hotspots, assess exposure risks, and develop targeted interventions to improve air quality and protect public health. IoT-based air quality measurement has applications across various sectors, including public health, environmental monitoring, urban planning, transportation, and industry. Notable applications include public health assessment, environmental monitoring, urban planning, transportation, and industrial facilities. However, challenges such as sensor accuracy, data privacy, data integration, and community equity must be addressed to realize its full potential. Despite these challenges, IoT technologies offer unprecedented opportunities to advance air quality measurement and monitoring, revolutionizing our understanding of pollution dynamics and informing evidence-based interventions to protect public health and the environment. Collaborative efforts from governments, industries, academia, and civil society are needed to leverage IoT innovations to build a cleaner, greener, and more prosperous future for all.

(Aswatha et al., 2023) suggested a smart air pollution monitoring system uses sensors to measure pollution levels and stores data in Firebase. The system forecasts air quality utilising machine learning methods. This technology is intended to provide real-time air quality monitoring. (Alaoui et al., 2019) Developed an accurate model for US air quality utilising data from wireless sensor networks (WSN), which was a key difficulty in predicting air pollution. The study investigates the possibility for using big data and the internet of things to anticipate air pollution caused by dangerous compounds such as NO_2, SO_2, CO, and O_3.

(Dhingra et al., 2019) The Internet of Things (IoT) is a global network of smart gadgets capable of sensing and interacting with their surroundings. To combat global air pollution, a three-phase air pollution monitoring system based on Arduino IDE, Wi-Fi modules, and gas sensors is proposed. The technology, which may be physically put in cities, accesses air quality data via an Android application called IoT-Mobair. The system can forecast pollution levels and future air

ACCEPTED MANUSCRIPT

quality index values. (Parmar et al., 2017) A working prototype of an environmental air pollution monitoring system has been developed by integrating Wi-Fi modules and inexpensive semiconductor gas sensors. The gadget uses a Raspberry Pi 3 server to display the data it collects on gas concentrations. The system's low-cost infrastructure for data delivery and collection is part of its design. (Gupta et al., 2019) suggested an IoT-based Air Quality Monitoring System for Smart Cities that gathers real-time data on temperature, humidity, carbon monoxide, LPG, smoke, and other harmful particulate matter from smart devices. The data is analysed and made available globally through an Android application, ensuring that cities stay livable and smart.

(Okokpujie et al., 2018) proposed an Internet of Things-based air quality monitoring system for smart cities that collects temperature data in real time. The data is analysed and made available globally through an Android application, ensuring that cities stay livable and smart. (Idrees & Zheng 2020) analysed air pollution monitoring systems (APMS), classifying them as static and mobile, portable devices, community-supported techniques, WSN-based systems, and IoT-supported approaches. It compares architecture and technologies, investigates real-time monitoring concerns, methods, and constraints, and proposes future goals for more precise and realistic air monitoring systems.

(Vallero , 2014) provided a better systems view and increased coverage of worldwide air pollution issues. Also covered were new materials on near-road air pollution, risk assessment methods, indoor air quality, biofuels, mercury emissions, forecasting tools, and the National Air Toxics Assessment. (Sierra-Vargas, 2012) analysed the relationship between air pollution and respiratory disease, proposes that lowering pollution can prevent disease, and emphasises the potential benefits of collaborative initiatives. (Naik et al., 2023) offered an Internet of Things (IoT)-based air pollution monitoring system to track pollution levels in industrial, residential, and

ACCEPTED MANUSCRIPT

transportation areas. The system can analyse pollutants, display pollution levels in any location, and save measured values in a cloud database.

(Senthilkumar et al., 2020) stated an air quality monitoring system based on fog computing and the IoT. The data generated by microprocessor-based IoT sensing devices is transferred to the cloud for rapid, high-volume service. The fog nodes filter non-actionable data before sending it to the cloud for long-term storage for batch analytics. This new technology for measuring air quality can detect changes in air quality patterns. (Yavas et al., 2021) investigated the impact of the Sakarya fireworks factory explosion on air pollution, highlighting the potential health effects of such events and the need for future research.

(Hardini et al ., 2023) considered Machine learning (ML) a powerful tool for developing an Air Quality Index (AQI). It involves collecting and preprocessing air quality data from sources like sensors, satellites, and weather stations. Feature engineering is then done to preprocess the data, ensuring consistency and improving model performance. ML algorithms are chosen based on the problem, data characteristics, and performance requirements. The dataset is split into training, validation, and test sets, and the models are trained and evaluated. The ML models can predict AQI values in real-time or on-demand, enabling continuous monitoring and timely updates. They are also adaptable and scalable, making them suitable for processing large volumes of data from multiple sources. However, careful consideration of data quality, model interpretability, bias, and ethical concerns is crucial for developing robust and reliable AQI systems.

(Song er al., 2017) looked into the impact of pyrotechnics on air pollution levels in a valley community. The results revealed that pollutant concentrations increased dramatically within 2-4 hours of the displays, with the highest concentrations reported in dwelling areas. However, dwelling quarters and industrial zones had lower SO_2 and NO_2 levels than preliminary Eve

ACCEPTED MANUSCRIPT

values. O3 concentrations dropped rather than rising with the displays. According to the study, interactions between fireworks and human activities lead to diverse pollution trends. (Chen et al., 2022) examined how fireworks regulations are implemented in China and found that strict regulations significantly improve public health by lowering rates of breathing and heart-related diseases, while moderate regulations have no effect and reduce ambient PM2.5 concentrations by 8% during festival months. These findings offer new insights for policymakers. For, instance, (Lai & Brimblecombe, 2020) Fireworks at China's New Year celebrations emit large levels of airborne particulate matter, necessitating tight regulations. Despite initial ineffectiveness due to mercantile and stakeholder pressures, advances have been observed in recent years, particularly in Beijing's inner districts. Sales of fireworks have fallen, as has debris accumulation. Consumer opinions have turned towards regulation, with social media and government articles emphasising the importance of safety when using fireworks. The strategy has proven effective in decreasing pollution from Beijing fireworks.

During firecrackers, optimization problems can arise, particularly during events like cultural festivals. These problems involve minimizing the negative impact of fireworks on air quality, public health, safety, and environmental sustainability. To address these issues, optimize the timing, location, and intensity of fireworks displays, choose environmentally friendly fireworks, and consider meteorological conditions. Control noise pollution by selecting low-noise fireworks, implementing noise abatement measures, and establishing buffer zones between fireworks sites and residential areas. Implement fire safety planning to prevent fires caused by fireworks, and prioritize public health protection for vulnerable populations. Optimize environmental conservation efforts to minimize the ecological impact of fireworks on wildlife, vegetation, and ecosystems. Optimize traffic flow and crowd control to ensure safe movement.

ACCEPTED MANUSCRIPT

Minimize littering and environmental contamination from fireworks debris and packaging materials. Foster community engagement and stakeholder collaboration to promote inclusivity and transparency in decision-making processes.

In this research IoT and Spark models can offer innovative solutions to monitor, analyze, and improve air quality. The main problems include limited monitoring infrastructure, health impacts, environmental degradation, economic costs, and regulatory compliance. Feasible solutions include IoT sensor networks for real-time data collection across various locations, data aggregation and analysis using Spark, and Air Quality Index (AQI) calculation using algorithms and models. Predictive modeling can forecast future air quality conditions, enabling proactive measures to mitigate pollution and protect public health. Alert systems based on AQI thresholds can notify authorities and the public about deteriorating air quality conditions, triggering appropriate responses such as advisories, restrictions, or interventions. Major contributions of IoT and Spark include real-time monitoring, data-driven insights, public awareness, policy formulation, and environmental impact. Real-time monitoring provides timely insights into pollution levels and trends, while data-driven insights facilitate data-driven decision-making for policymakers, urban planners, and environmental agencies. The AQI raises awareness about the importance of reducing pollution and taking preventive measures. Accurate air quality data generated by IoT and Spark models can inform the development of evidence-based policies and regulations aimed at improving air quality and protecting public health. Therefore, integrating IoT sensors with Spark-based analytics to calculate an Air Quality Index offers a powerful solution to address air quality challenges, providing real-time monitoring, data-driven insights, and actionable information for stakeholders to improve public health and environmental quality.

2. Materials and methods

ACCEPTED MANUSCRIPT

The suggested system design for monitoring air quality in fire work industrial settings is explained in this section.

2.1. Architecture design:

The main layer and information source of the Internet of Things is the perception layer. This layer uses sensors, electronic data interfaces (EDI), radio frequency identification (RFID) systems, wireless sensor networks (WSN), global positioning systems (GPS), reader-writers, objects, tags, smart terminals, cameras, and other technologies to recognise objects and collect data. Due to the integration of many technologies, such as WSNs, RFID sensor networks (RSN), and RFID, the idea of the Internet of Things has changed over time. A reader and a few tags are the foundation of an RFID system, which uses radio waves to read and record data stored on tags attached to objects. RFID can identify things wirelessly and without requiring a direct line of sight. Sensing and monitoring the environment are the goals of WSN. Figure 2 displays the architecture design. The system operates on its own for varying durations, ranging from days to years. Numerous sensor nodes that can be employed on the ground, in the air, in vehicles, within buildings, etc. make up the sensor network. RSN is an RFID and WSN integration that helps locate and identify objects while also giving information about their state to the owner of the sensor-enabled RFID tag.

ACCEPTED MANUSCRIPT

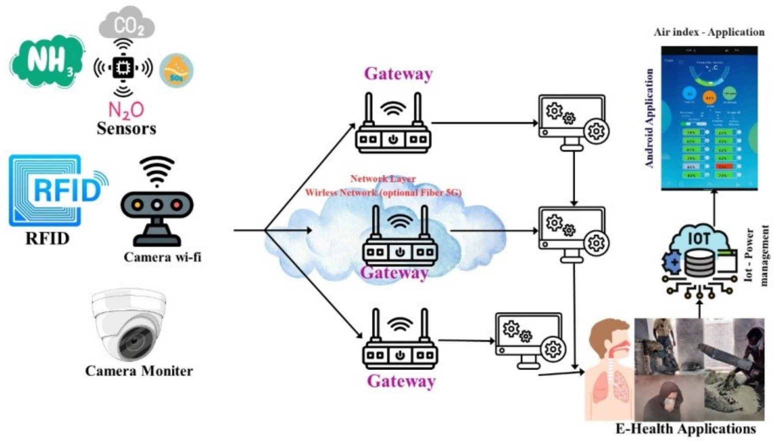


Figure 2 Architecture design

The novelty of using the Internet of Things (IoT) to create an Air Quality Index (AQI) stems from its capacity to improve accessibility, data accuracy, and real-time monitoring.

2.2. Integration of IoT algorithm for fireworkers

Fast and versatile cluster computing solution, Apache Spark is also an open-source processing platform. Figure 3 shows the ecosystem of spark. It facilitates the quick construction of big data applications; it allows programming languages with high-level APIs, including Scala, Java, R, and Python; and it has an optimised engine that can handle general execution graphs. It also offers an extensive array of advanced devices, such as Spark SQL for SQL, GraphX for graph processing, Spark streaming, and MLlib for ML, as well as structured data analysis.

ACCEPTED MANUSCRIPT

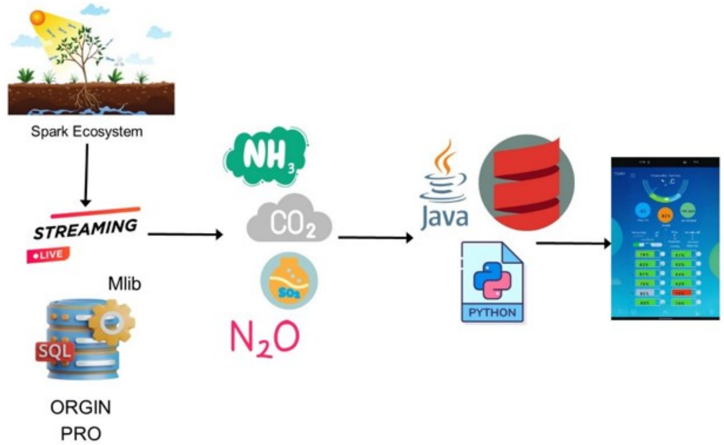


Figure 3 Spark network

Spark models are a distributed computing framework that offers scalability, parallel processing, real-time analytics, and ecosystem integration for developing Air Quality Index (AQI) prediction systems. They handle large-scale data sets from multiple sources, enabling efficient data processing and analysis. Spark's machine learning library (MLlib) provides a rich set of scalable algorithms and tools for building predictive models. Real-time analytics through Spark Streaming and Spark SQL enable AQI prediction models to analyze streaming data streams and generate real-time predictions. However, Spark models require expertise in distributed computing, data engineering, and machine learning, requiring additional complexity and overhead. Resource constraints may also be a challenge, with Spark models requiring significant computational resources. Algorithm selection depends on data characteristics, modeling objectives, and performance requirements. Data quality and preprocessing are crucial for accuracy and reliability. Interpretability and transparency are essential for building trust and validating results. In conclusion, Spark models' effectiveness depends on various factors, and careful evaluation of problem requirements, computational resources, data characteristics, and modeling objectives is necessary to determine their suitability for AQI prediction.

ACCEPTED MANUSCRIPT

Simply put, Spark M_{Llib} is Spark's ML library. It provides a wide range of capabilities, including pipelines, ML algorithms, persistence, and feature optimisation. Figure 4 shows the algorithm outline. It seeks to simplify and scale practical machine learning. In this study, the investigators had selected:

• Gradient-boosted trees (GBTs): a widely used regression method that makes use of ensembles of decision trees, in addition to an effective classification procedure. Decision trees are iteratively trained by GBTs to reduce a loss function. Similar to decision trees, GBTs may record feature interactions and nonlinearities, work in classification with multiple classes settings, and handle categorical information. They also don't require feature scaling.

• ML pipelines: a group of tools that let users build, assess, and fine-tune ML pipelines.

This work used an easy-to-understand and efficient method to analyse our data and build a model centred on the GBT classifier.

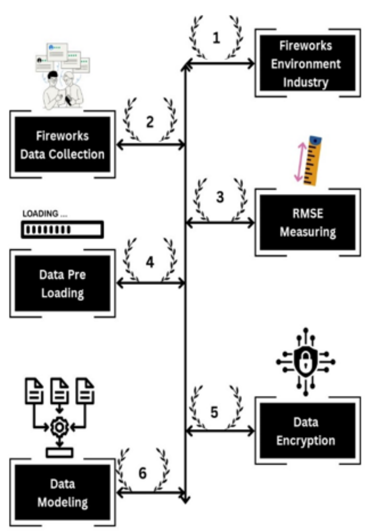


Figure 4 Outline of the algorithm

ACCEPTED MANUSCRIPT

Step 1: Preparing the data

It entails preparing data so that it may be processed by the user more quickly and efficiently because most data is raucous, erratic, and imprecise.

Step 2: Configuring the environment

This part set up an account on Databricks (no date), a unified analytics system built by the people who developed Apache Spark with the intention of assisting customers with large data processing on the cloud using Spark.

Step 3: Loading data

The process of transferring and loading a dataset into Databricks' distributed file system, DBFS, from its source file is known as data loading.

Step 4: Dividing data

The technique of separating accessible data into two halves, usually for cross-validation, is known as data splitting. A piece of the data is used to develop a predictive model (called training data), while another portion is used to assess the model's efficacy (called test data).

Step 5: Modelling data

Data modelling is the process of creating an ML model to forecast future air quality and offer advice on the health effects of air pollution on people.

Step 6: Measure the Root Mean Square Error (RMSE)

RMSE is a well recognised metric that quantifies the discrepancies between the results that the model predicts and the actual values that are observed.

ACCEPTED MANUSCRIPT

2.3. Measuring the Air Quality Index of fireworkers

2.3.1. Prediction

The local weather and neighbouring emissions of pollutants have a substantial correlation with air quality levels. Nevertheless, when predicting local AQI levels, long-range pollution transport—via high winds—must be taken into account as a significant influencing factor. As a result, forecasting air quality entails more than just the challenges associated with weather prediction; it also necessitates data and knowledge regarding both local and distant pollution levels and discharges. Figure 5 shows the overall flow of the suggested system.

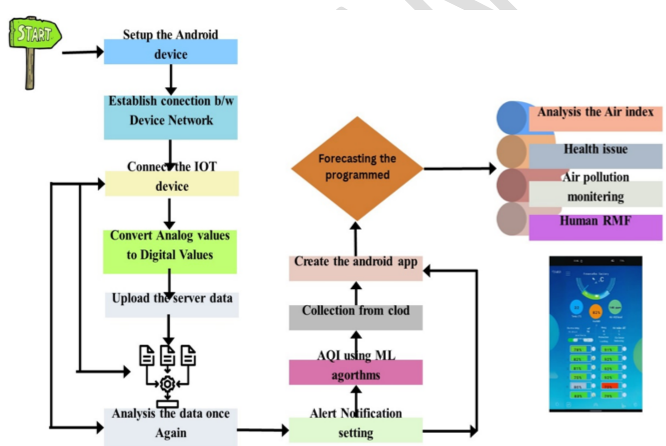


Figure 5 Prediction of the proposed system

2.3.2. Calculating the health condition of fireworkers through AQI

One of the key instruments for evaluating and consistently portraying the state of the air quality is the AQI. The Air Quality Index (AQI) is a common way to describe the total effect of individual pollutant concentrations in the surrounding air as a single number which is shown in table 1. The following formula has been used to calculate the AQI index for the data set used for

ACCEPTED MANUSCRIPT

training.

Table 1. Health Breaks, Pollutants, and AQI Category

AQI Category (Range)	Good (0-50)	Satisfactory (51-100)	Moderately polluted (101-200)	Poor (201-350)	Very poor (301-400)
PM_{10}	0-40	31-90	71-240	115-250	251-350
PM_{25}	0-20	21-50	41-80	71-110	111-150
NO_2	0-30	31-80	91-170	161-250	271-648
O	0-40	41-90	91-158	159-108	109-548
CO_3	0-1.5	1-2.5	1.1-15	15-27	17-64
SO_2	0-30	31-70	61-280	281-750	701-1500

2.5. Android application for fire workers to calculate AQI

Using regression models like GBT and Spark MLlib and functions from Python machine learning libraries, the model forecasts the accuracy of the air purity index level in years to come. There is a comparison between the two models' accuracy results which is shown in figure 6.

ACCEPTED MANUSCRIPT

Figure 6 The Android application's dashboard screen

The standard pre-script denotes the standard value in accordance with Central Pollution Control Board (CPCB) regulations, while the actual pre-script shows the observed value of the parameter for a given duration. As seen in figure 7, a smartphone app is created to show the current AQI number, the current temperature and humidity readings, and the breakdown of all the gases that cause pollution. Additionally, the app notifies the user in the event of extreme air pollution and shows the next AQI as anticipated by the machine learning model.

ACCEPTED MANUSCRIPT

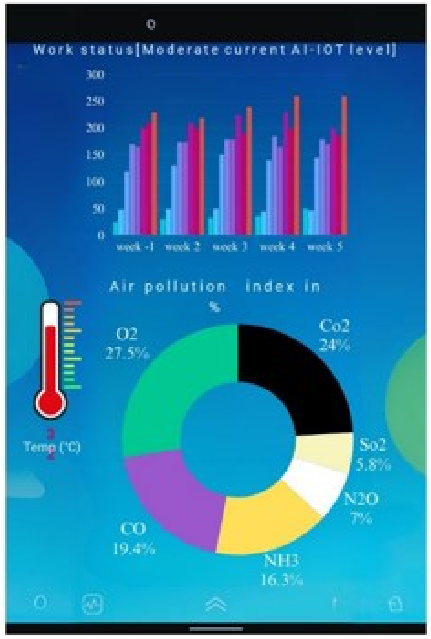


Figure 7. The Android application's forecast screen

3. Results and discussion

3.1. Collection of data

To lower the hazards, governments and organisations have imposed regulatory limitations on certain pollutants. Table 2 displays the various standard limits and pollution levels that have been declared. These details were gathered from **Sri Krishna Fireworks in Tamil Nadu's Sivakasi**, also known as "mini Japan." For this study, crackers such as lanterns, rockets, atom bombs, and crackling king were employed. The IoT sensor used to detect air quality index they are temperature sensor, humidity sensor, proximity sensor, vibration sensor, pressure sensor and dust sensor which is employed in table 2.

Table 2 Data collection

ACCEPTED MANUSCRIPT

Pollutant	IoT sensor used	AIR QUALITY INDEX (AQHI LEVEL) taken in sivakasi (Sri Krishna Fireworks)				
		Sparklers	**Rocket**	**Atom bomb**	**lanterns**	**Crackling king**
Carbon Monoxide (CO)	Temperature sensor, Humidity sensor, Pressure sensor, Proximity sensor, Vibration sensor.PPD42 Dust Sensors	7.78ppm	13.23ppm	3.12ppm	3.89ppm	8.12ppm
Lead (Pb)		8.56 $\mu g/m^3$	16.89 $\mu g/m^3$	2.12 $\mu g/m^3$	4.23 $\mu g/m^3$	3.09 $\mu g/m^3$
Nitrogen Dioxide (NO_2)		7.56ppb	23.90ppb	1.23ppb	9.23ppb	2.34ppb
Ozone (O_3)		7.12 $\mu g/m^3$	34. $\mu g/m^3$	2.12 $\mu g/m^3$	17.98 $\mu g/m^3$	18.90 $\mu g/m^3$
Particulate Matter ($PM_{2.5}$ and PM_{10})		3.89 $\mu g/m^3$	23.12 $\mu g/m^3$	3.12 $\mu g/m^3$	8.23 $\mu g/m^3$	23.12 $\mu g/m^3$
Sulfur Dioxide (SO_2)		8.89ppb	32.22ppb	1.23ppb	6.98ppb	31.12ppb

ppm-part of gas per million in air $\mu g/m^3$- cubic meter of air contains one microgram ppb-part of gas per billion in air

It is a crucial tool for predicting air quality, requiring a thorough analysis of factors like pollutant concentrations, meteorological conditions, and historical data. Machine learning (ML) techniques can be used to predict AQI by analyzing these data. The process involves data collection, preprocessing, model selection, training, evaluation, deployment, and continuous improvement. Data is collected from various sources, including air quality monitoring stations, satellite observations, and weather stations. ML algorithms are chosen based on the dataset size, complexity, and desired accuracy. The model is trained using the training dataset, optimized

ACCEPTED MANUSCRIPT

through techniques like cross-validation, and evaluated using the testing dataset. Continuous improvement is encouraged through regular updates and the incorporation of new data and features. The government and organisation agencies created an indicator called the Air Quality Index (AQI) to make it easier for the public to comprehend the current state of the air quality. When it comes to meeting the needs of one or more biotic species, as well as any human need or purpose, the AQI gauges each person's "condition or state." Put simply, it informs the public of how "good" the air quality is right now or is expected to be in the future. Air quality indexes may vary throughout agencies.

3.2. Data Measurement of Gas in air using data samples

The researcher conducted a study on the occupational safety and well-being of Sivakasi's fireworks workers. Using basic random selection, researchers chose 10 samples—six male and four female—from among those who have worked with fireworks at Sri Krishna Fireworks for their study. A survey of fireworks industry personnel was undertaken by a researcher to obtain data on the efficacy of the industry. IoT sensors were employed by the researcher as a data collection tool. Table 3 displays the measurements made by the polluting or contaminating gas measuring device.

Table 3 Data Measurement scale

Statistics	No of Observations (6 male and 4 female)	Average value Min	Average value Max	Range	Standard deviation(σ)	Air Quality Index
CO	10	2.2	3.5	0.453	0.5637	4.213

ACCEPTED MANUSCRIPT

Particulate Matter ($PM_{2.5}$ and PM_{10})	10	15.32	17.90	13.28	22.994	20.89
Lead (Pb)	10	19.23	20.21	15.32	3.7822	16.90
SO_2	10	24.9	26.3	4.21	4.9082	16.23
O_3	10	0.43	4.21	3.90	5.9073	4.90
NO_2	10	0.23	6.21	2.34	2.0001	3.90

The Air Quality Health Index (AQHI) system, which was created by the Pollution Control Board of India, is used as an example of an AQI to help explain the concept. The figure 8 shows this system. The AQHI method offers a clearer picture of the public's health hazards and offers specific recommendations for preventative measures for each AHQI level.

Figure 8 AQHI level

The following outcomes were attained after researchers used Python in Spark to create our model, which was built on GBTs and ML pipelines: Our data is first transformed into a

ACCEPTED MANUSCRIPT

DataFrame, or df, that is a distributed set of data arranged into columns with names. The count() function was then utilised, which yields the total number of components in a set of data or the outcome of an RDD operation. Upon launch, the Android app requests authentication. To log in to the application, the user can enter a password and a username. AQI forecast and Dashboard are the 3 navigation boards in the user interface. The Android application's dashboard, shown in Figure 9, shows the current AQI level in parts per million (ppm) as well as the makeup of each pollutant, including Pb, Particulate matter, Ozone, NO_2,CO, and SO_2. The dashboard additionally shows the humidity and temperature. The training data sets' numerical and visual representations can be found in the Android application's graphs navigation, which is used to analyse real-time data values. This study confined the columns shown to:

1. The actual value of the air's purity in order to make the results simpler to see, as in Figure 10.
2. Based on our model, the expected air quality rating.
3. The remaining feature columns in our dataset are called featurecols.

ACCEPTED MANUSCRIPT

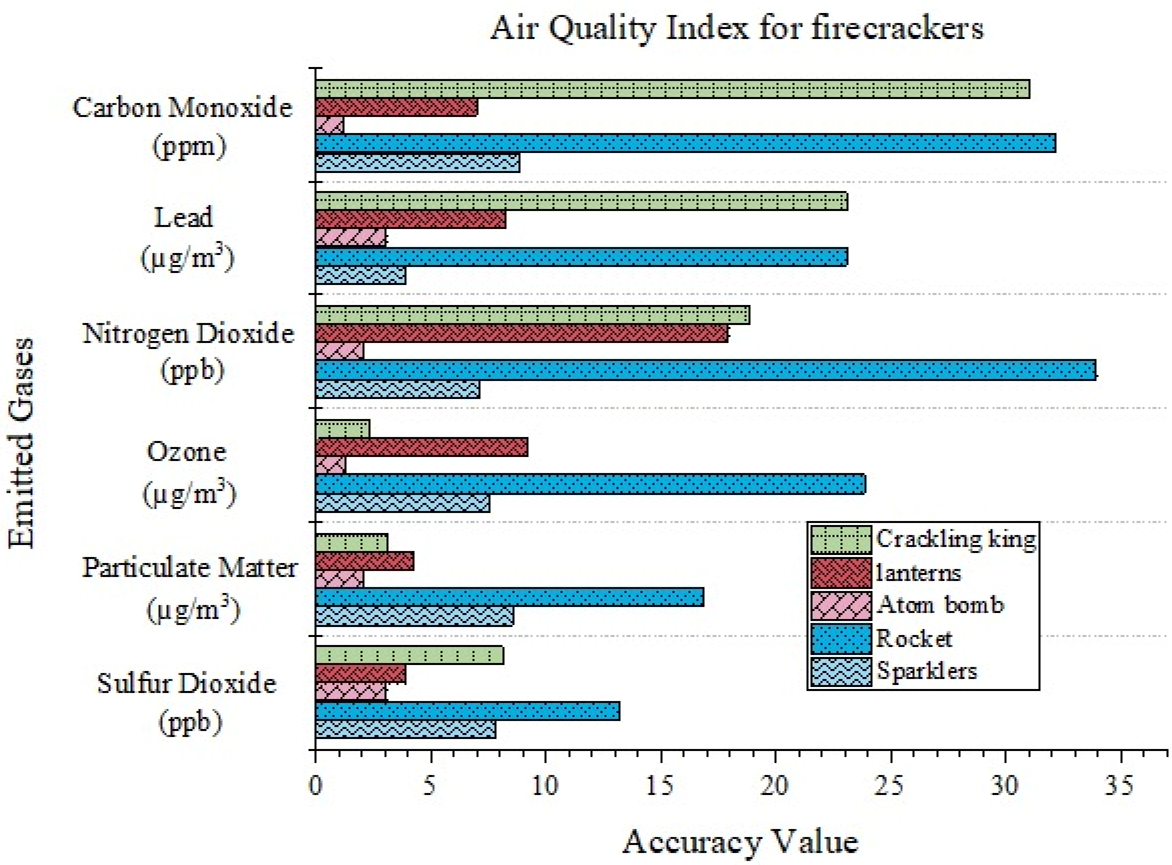


Figure 9 Accuracy value of emitted gases in atmosphere during blasting firecrackers

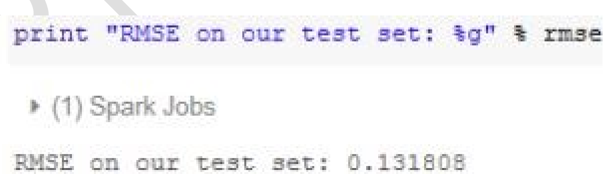


Figure 10 RMSE value

ACCEPTED

The AQI Forecast navigator of the application for Android, depicted in Figure 10, shows the expected AQI of the upcoming days together with the AQI level chart. Lastly, calculating assessment measures is critical to comprehending forecast quality. RMSE was employed in this study to assess our predictions. RMSE values for good prediction models should generally be less than 0.3. It is true that our RMSE (Figure 10) equals 0.14 < 0.4, which leads us to conclude that our model is reliable.

7. Conclusion

This paper provides comprehensive details on the design and implementation of a smart air pollution tracking system developed by researchers. The system continuously monitors the quality of the air in a given area, analyses the data using a training data set, and forecasts the pollution level for the next few days. This work has focused on combining two novel ideas—big data and IoT—to address environmental problems, including air pollution. In fact, by utilising Spark technology in the Databricks structure and a US pollution dataset, this study was able to construct a precise model that can foresee air quality well enough to aid in our understanding of the detrimental effects that air pollution has on our lives and in our efforts to avoid, manage, and minimize this problem as soon as feasible. During the Diwali festival, loud explosive fireworks displays, cracker displays, etc., produce significant but transient air pollution. The amount of SO_2 in firecrackers rose many times, and the amount of air pollutants doubled compared to what is normally reported on a regular winter day. The area with the greatest amount of air pollutants was identified as S15, which appears to be associated with the high population density and economic standing of the local residents. Technology plays a crucial role in evaluating air pollution indexes (APIs) by providing accurate and timely data collection, analysis, and dissemination. Advanced air quality sensors, such as IoT-based sensors, provide real-time monitoring of pollutants like PM, NO_2, SO_2, CO, and O_3. Remote sensing technologies, such as satellite-based sensors, offer a broader perspective on air pollution. Big data analytics and

ACCEPTED MANUSCRIPT

machine learning enable the processing and analysis of large volumes of data, generating insights for API evaluation. Visualization tools and geographic information systems help stakeholders understand API values, pollution trends, and potential health impacts. Mobile applications provide personalized API information, while integrated technology platforms support policymakers in formulating and implementing air quality policies. Future work can include applying the technique of forecasting to bigger areas with more intricate machine learning algorithms.

References

Aswatha, S., R. Deepika, M. Dharu Piraba, V. P. Dhaneesh, K. Madheswari, S. Saraswathi, Y. V. Lokeswari, and Nagarajan KK. (2023)"Smart air pollution monitoring system." *GLOBAL NEST JOURNAL* **25**, no. 3: 125-129.

Alaoui, Safae Sossi, Brahim Aksasse, and Yousef Farhaoui. (2019) "Air pollution prediction through internet of things technology and big data analytics." *International Journal of Computational Intelligence Studies* **8**, no. 3: 177-191.

Dhingra, S., Madda, R. B., Gandomi, A. H., Patan, R., & Daneshmand, M. (2019). Internet of Things mobile–air pollution monitoring system (IoT-Mobair). *IEEE Internet of Things Journal*, ***6***(3), 5577-5584.

Parmar, G., Lakhani, S., & Chattopadhyay, M. K. (2017, October). An IoT based low cost air pollution monitoring system. In *2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)* (pp. 524-528). IEEE.

Gupta, H., Bhardwaj, D., Agrawal, H., Tikkiwal, V. A., & Kumar, A. (2019, February). An IoT based air pollution monitoring system for smart cities. In *2019 IEEE International*

Conference on Sustainable Energy Technologies and Systems (ICSETS) (pp. 173-177). IEEE.

Hardini, M., Sunarjo, R. A., Asfi, M., Chakim, M. H. R., & Sanjaya, Y. P. A. (2023). Predicting air quality index using ensemble machine learning. *ADI Journal on Recent Innovation*, *5*(1Sp), 78-86.

Okokpujie, K., Noma-Osaghae, E., Modupe, O., John, S., & Oluwatosin, O. (2018). A smart air pollution monitoring system. *International Journal of Civil Engineering and Technology (IJCIET)*, *9*(9), 799-809.

Idrees, Z., & Zheng, L. (2020). Low cost air pollution monitoring systems: A review of protocols and enabling technologies. *Journal of Industrial Information Integration*, ***17***, 100123.

Vallero, D. A. (2014). *Fundamentals of air pollution*. Academic press.

SIERRA-VARGAS, M. P., & Teran, L. M. (2012). Air pollution: Impact and prevention. *Respirology*, ***17*(**7), 1031-1038.

Naik, U. U., Salgaokar, S. R., & Jambhale, S. (2023). IOT based air pollution monitoring system. *Int. J. Sci. Res. Eng. Trends*, ***9***, 835-838.

Senthilkumar, R., Venkatakrishnan, P., & Balaji, N. (2020). Intelligent based novel embedded system based IoT enabled air pollution monitoring system. *Microprocessors and Microsystems*, *77*, 103172.

Yavaş, S. P., Baysan, C., & Önal, A. E. (2021). The Effect of Firework Explosion at the Fireworks Factory on Air Pollutant Levels. *The Anatolian Journal of Family Medicine*, ***4*(**1), 80.

ACCEPTED MANUSCRIPT

Song, Y., Wan, X., Bai, S., Guo, D., Ren, C., Zeng, Y., ... & Li, X. (2017). The characteristics of air pollutants during two distinct episodes of fireworks burning in a Valley City of North China. *PloS one*, ***12***(1), e0168297.

Chen, S., Jiang, L., Liu, W., & Song, H. (2022). Fireworks regulation, air pollution, and public health: Evidence from China. *Regional Science and Urban Economics*, ***92*,** 103722.

Lai, Y., & Brimblecombe, P. (2020). Changes in air pollution and attitude to fireworks in Beijing. *Atmospheric environment*, ***231***, 117549.

ACCEPTED MANUSCRIPT