International Journal of Research and Review in Applied Science, Humanities, and Technology
Vol 1, Issue 2 December 2024 ISSN: 3048-975X
https://ijrasht.com/

Optimized Air Quality Index and Meteorological Predictions with Machine Learning and IoT

Dr. Venkateswaran Radhakrishnan, Information Technology Department, College of Computing and Information Sciences, University of Technology and Applied Sciences, Salalah, Sultanate of Oman Venkateswaran.Radhakrishnan@utas.edu.om

Dr. Suresh Palarimath, Information Technology Department, College of Computing and Information Sciences, University of Technology and Applied Sciences, Salalah, Sultanate of Oman Suresh.palarimath@utas.edu.om

Mr. Rogelio Gutierrez, Information Technology Department, College of Computing and Information Sciences, University of Technology and Applied Sciences, Salalah, Sultanate of Oman Rogelio.Gutierrez@utas.edu.om

Abstract: Air Quality Index (AQI) prediction and forecasting play pivotal roles in assessing and managing air pollution, contributing to public health and environmental sustainability. This paper provides a comprehensive review of recent advancements, methodologies, challenges, and future directions in AQI prediction and forecasting. Recent research has seen a surge in the development of machine learning, statistical, and hybrid models for AQI prediction. These models leverage various input data sources such as meteorological data, satellite imagery, and pollutant emissions data to enhance prediction accuracy. Furthermore, the integration of advanced techniques like deep learning and ensemble modeling has shown promising results in capturing complex nonlinear relationships and improving forecast precision. Challenges persist, including the need for real-time data integration, model interpretability, and addressing spatial and temporal variations in air quality. Additionally, the impact of emerging factors such as climate change and urbanization on AQI prediction requires further investigation. Future research directions focus on the development of hybrid models that integrate multiple data sources, including sensor networks and IoT devices, to improve spatial and temporal resolution. Moreover, there is a growing emphasis on the incorporation of uncertainty quantification techniques to provide probabilistic forecasts and enhance decision-making under uncertainty. In conclusion, this paper underscores the importance of AQI prediction and forecasting in addressing air pollution challenges and promoting public health. By advancing methodologies, addressing challenges, and exploring emerging research avenues, we can strive towards more accurate, reliable, and actionable AQI predictions for sustainable urban development and environmental stewardship.

Keywords: AQI, Machine Learning, Metrological Parameter, IoT, Hybrid

1. INTRODUCTION

Air quality is a critical component of environmental health, with profound implications for human well-being and ecosystem sustainability. The deterioration of air quality due to pollutants emitted from various sources poses significant challenges to public health and environmental management worldwide. The Air Quality Index (AQI) is a pivotal tool for evaluating and conveying air quality levels to individuals and policymakers. It provides a standardized metric that quantifies the concentration of key air pollutants and their potential health effects, thereby aiding in decision-making processes related to pollution control and public health protection. PM2.5, PM10, O3, CO, NO2, and SO2 are six aerosols that are significant in judging the air quality index [1]. Over the years, the measurement and prediction of AQI have garnered increasing attention from researchers, environmental agencies, and policymakers. The importance of accurate and timely AQI information cannot be overstated, as it enables stakeholders to monitor air quality trends, identify pollution hotspots, and implement targeted interventions to mitigate adverse effects on human health and the environment. Traditional methods of AQI measurement primarily rely on ground-based monitoring stations equipped with sophisticated instrumentation to measure pollutant concentrations at specific locations. While these stations offer reliable data, their spatial coverage may be limited, resulting in gaps in monitoring networks, especially in remote or underdeveloped regions. In recent years, technological advancements have revolutionized AQI measurement and prediction capabilities, ushering in a new era of data-driven approaches and innovative methodologies. The proliferation of low-cost air quality sensors, coupled with advancements in data analytics and machine learning techniques, has facilitated the development of high-resolution AQI maps and real-time monitoring systems. These advancements not only enhance spatial coverage but also improve the temporal resolution of AQI data, enabling stakeholders to obtain up-to-date information on air quality fluctuations and trends. [2,3 ,4]

2. BACKGROUND SURVEY

The literature survey/background encompasses key studies that offer insights into air quality assessment and prediction, thereby informing the objectives and methodology of the "Cutting-edge Weather Station: Advanced Monitoring of AQI and Meteorological Forecasting" project. Firstly, the research conducted by Ankita P. Dadhich et al. (2017), titled "Assessment of spatio-temporal variations in air quality of Jaipur city, Rajasthan, India," provides a relevant assessment of air quality dynamics in an urban environment. Employing geospatial and geostatistical techniques, the study evaluates seasonal variations in pollutants, identifies major contributors to air degradation, and correlates air quality with local weather parameters. This investigation underscores the critical importance of spatial analysis in identifying pollution hotspots and guiding effective mitigation strategies [5]. Furthermore, the work of M. Pulikesi et al. (2006), "Air quality monitoring in Chennai, India, in the summer of 2005," offers valuable insights into surface ozone (O3) dynamics and its correlation with meteorological parameters and anthropogenic influences. The study underscores the urgency for targeted pollution control measures in urban environments, particularly in mitigating Total Suspended Particulate Matter (TSPM) levels exceeding National Ambient Air Quality Standards (NAAQS) [6].

Additionally, the research by Anikender Kumar and P. Goyal (2011) on "Forecasting of daily air quality index in Delhi" employs statistical methods to develop forecasting models for daily Air Quality Index (AQI) in Delhi. This study highlights the significance of meteorological parameters in predicting AQI levels, emphasizing the need for accurate forecasting models to support public health management and urban planning initiatives [7]. Finally, the recent work by R. Janarthanan et al. (2021) titled "A deep learning approach for prediction of air quality index in a metropolitan city" explores advanced methodologies, including deep learning and statistical models, for air quality prediction in urban settings [8]. This study underscores the critical role of accurate prediction in addressing contemporary challenges in public health management and urban planning, thus aligning with the objectives of the "Cutting-edge Weather Station" project. Through these relevant literature surveys, the project gains valuable insights into air quality assessment, prediction methodologies, and the interaction of meteorological factors and air pollution, ultimately aiding in research knowledge and decision-making in environmental management and public health oversight.

2.1 Previous & Related Work

Previous work in the field of weather monitoring and air quality assessment has laid the foundation for this project by exploring various methodologies and technologies for data collection and analysis. Traditional weather monitoring methods, such as manual observations and basic weather stations, have been widely used but are limited in their spatial coverage and data accuracy. Recent advancements in technology, including the integration of IoT devices and remote sensing techniques, have led to the development of more sophisticated weather monitoring systems capable of providing real-time data with higher precision. Similarly, in the realm of air quality assessment, previous studies have focused on the correlation between air pollutants and meteorological factors, highlighting the importance of monitoring both parameters simultaneously for accurate forecasting and risk assessment. Additionally, by deploying networks of sensors utilizing satellite-based evidence, research has been done to broaden the dynamic range and geographical accuracy of air pollution tracking networks. By building upon the findings of previous research and leveraging emerging technologies, this project aims to develop a cutting-edge weather station capable of advanced monitoring of air quality index (AQI) and meteorological parameters for enhanced forecasting and decision-making in various sectors [9,10].

2.2 Machine Learning and Algorithm

Support Vector Machines (SVM) are a type of supervised learning model used for classification and regression tasks. Support vector machines (SVMs) are models for supervised learning that address data utilized for regression and classification. They are paired with learning algorithms. [11,12]. The key idea behind SVM is to find the hyperplane that optimally divides the classes in the feature space. This hyperplane is chosen in such a way that it maximizes the margin, which is the distance between the hyperplane and the nearest data point from each class, also known as support vectors. SVMs are effective in high-dimensional spaces and are particularly well-suited for cases where the number of dimensions exceeds the number of samples. They are also versatile because different kernel functions can be specified for the decision function. Common kernel functions are linear, polynomial, radial basis function (RBF), and sigmoid. Training an SVM involves finding the optimal hyperplane that separates the classes. This is done by solving a convex optimization problem, typically using techniques like gradient descent or quadratic programming. After being trained, the SVM can recognize new data points by recognizing the edge of the hyperplane they fall on.

Random Forest (RF) Within the collaborative training domain, Random Forest is an appreciated machine learning technique. It is a collection of decision trees, where each tree is trained on a random subset of the training data (bootstrapped samples) and at each split in the tree, only a subgroup of the features is acknowledged. The RF approach uses triggering classification and regression trees (CARTs). Each CART is built up of random vectors. [13]. Random Forest is an ensemble learning approach, which means it combines the predictions of numerous independent models to get a single final prediction. Random Forest's individual models are decision trees.

Multivariate linear regression is a variant of standard linear regression that anticipates a dependent variable (such as AQI) using countless independent variables (features or predictors). In the context of AQI prediction, multivariate linear regression considers not just one, but several factors that influence air quality. A multiple linear regression model is predictions utilized to explain a correlation between one continuous dependent variable and two or more independent variables.[11]. The multivariate linear regression equation extends the simple linear regression equation by accounting for numerous independent variables: [14]

$AQI = \beta 0 + \beta 1*X1 + \beta 2*X2 + ... + \beta n*Xn + \varepsilon$
AQI is the predicted Air Quality Index.
$\beta 0, \beta 1, \beta 2, ..., \beta n$ are the coefficients of the linear equation.
X1, X2, ..., Xn are the independent variables (features).
The error term, ε, reflects the disparity between the predicted and actual AQI.

XGBoost It can effectively predict Air Quality Index (AQI) by leveraging its ensemble of decision trees and gradient boosting techniques, offering accurate forecasts based on historical data of environmental factors and pollutant concentrations. Random Forest (RF) and XGBoost are decision tree-based models that vary in that XGBoost can decrease mistakes whereas RF cannot. XGBoost belongs to the ensemble learning family, specifically boosting algorithms. XGBoost utilizes the gradient boosting framework, which minimizes a predefined loss function by adding weak learners (decision trees) to the ensemble. It optimizes the model's performance by iteratively improving upon the residuals of the previous models. It also allows users to control the complexity of individual trees and the overall model through hyperparameters. It utilizes advanced techniques like parallelization, approximate tree learning, and tree pruning to make training faster and more efficient, even for large datasets. Overall, XGBoost is a versatile and reliable algorithm that excels in predictive modelling tasks.

KNN Using k-Nearest Neighbors (kNN) for Air Quality Index (AQI) prediction involves a straightforward approach where you predict the AQI for a given location and time based on the AQI values of its nearest neighbors. Here is how it can be applied in a sentence, k-Nearest Neighbors (kNN) can predict Air Quality Index (AQI) by finding the AQI values of the nearest neighboring locations and averaging or weighting them to estimate the AQI for a specific location, making it a simple yet effective method for spatial AQI prediction. Growing info leads to KNN to slow down, resulting in its main problem. By utilizing, one may calculate the Euclidean distance (d) among the two points (a1 − a2) and (b1 − b2).

$$\mathbf{d = ((a1 - b1)^2 + (a2 - b2)^2)^{1/2}}$$

3. PROPOSED METHODOLOGY

Data Collection: The weather station collects real-time data on various meteorological parameters, including temperature, humidity, wind speed, wind direction, atmospheric pressure, and precipitation. Particulate matter (PM), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), carbon monoxide (CO), and VOCs (volatile organic compounds) include various other air quality indicators that it examines.

Sensor Integration: The station integrates a variety of sensors to capture data from different environmental sources. These sensors may include temperature sensors, humidity sensors, anemometers, barometers, rain gauges, air quality sensors (e.g., PM sensors, gas sensors), and remote sensing devices such as cameras or lidar systems for additional data collection.

Data Transmission: The collected data is transmitted to a central processing unit or data logger within the weather station. Depending on the setup, this data transmission may occur through wired connections or wireless communication technologies such as Wi-Fi, cellular networks, or satellite communication.

Data Processing and Analysis: Once the data is received, it undergoes processing and analysis to extract meaningful insights. Machine learning algorithms may be employed to analyze historical data, identify patterns and trends, and develop predictive models for forecasting weather conditions and air quality levels. This analysis may also involve feature extraction, anomaly detection, and quality control procedures to ensure data integrity.

Visualization and Reporting: The processed data is then presented in a user-friendly format for visualization and interpretation. This may include graphical displays, charts, maps, and dashboards that provide real-time updates on current weather conditions, air quality levels, and forecasted trends. Users can access this information through web-based interfaces, mobile applications, or other communication channels.

Decision Support and Alerts: The weather station may incorporate decision support systems to provide actionable insights and alerts based on predefined thresholds or criteria. For example, it can issue alerts for extreme weather events, high pollution levels, or other hazardous conditions, enabling stakeholders to take timely actions to mitigate risks and protect public health and safety.

3.1 Dataset Description

Data Sources: Describe the sources of data utilized in the project, which may include:
- Meteorological Data: Obtained from weather stations, satellites, or other remote sensing devices. Includes parameters such as temperature, humidity, wind speed, wind direction, atmospheric pressure, and precipitation. [17]
- Air Quality Information: Obtained via sensors or stations that monitor air quality. comprised of gauges of several air pollutants such as carbon monoxide (CO), sulfur dioxide ($SO2$), nitrogen dioxide ($NO2$), particle matter (PM), ozone ($O3$), and volatile organic compounds (VOCs). [18]

Temporal Coverage: Specify the time covered by the dataset, including the frequency of data collection (e.g., hourly, daily, monthly).

Spatial Coverage: Outline the geographic extent of the dataset, including the locations of weather stations and air quality monitoring stations from which data was collected.

Data Format: Describe the format of the data, including file types (e.g., CSV, JSON) and any specific data structures or schemas used.

Data Variables: Provide a list of variables or parameters included in the dataset, along with their units of measurement and descriptions. This may include both meteorological parameters (e.g., temperature in degrees Celsius, humidity in percentage) and air quality indicators (e.g., PM2.5 concentration in micrograms per cubic meter, NO2 concentration in parts per billion).

Data Quality: Discuss any quality control measures implemented to ensure the accuracy and reliability of the data, such as calibration procedures, sensor maintenance, and outlier detection algorithms.

Data Preprocessing: Detail any preprocessing steps applied to the raw data before analysis, such as data cleaning, normalization, or feature engineering.

Data Availability: Specify the availability of the dataset for research purposes, including any restrictions on access or usage.

3.2 AQI Calculation with Data

Calculating the Air Quality Index (AQI) involves converting measured concentrations of air pollutants into a standardized index that reflects the relative level of air quality and associated health risks. Particulate matter (PM2.5 and PM10), ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2), and the amount of carbon monoxide (CO) ratios are some of the main pollutants that are routinely utilized in calculating the air quality index (AQI). [15,16]

Pollutant Concentrations: Obtain the measured concentrations of each air pollutant from the sensors installed in the weather station. These concentrations should be in units such as micrograms per cubic meter ($\mu g/m^3$) for particulate matter and parts per million (ppm) for gases like NO2, SO2, and CO.

Pollutant Categories and Breakpoints: Refer to the national or international standards for air quality, which define specific concentration ranges for each pollutant category (e.g., Good, Moderate, Unhealthy, etc.). These concentration ranges are referred to as breakpoints and vary depending on the pollutant.

International Journal of Research and Review in Applied Science, Humanities, and Technology
Vol 1, Issue 2 December 2024 ISSN: 3048-975X
https://ijrasht.com/

AQI Calculation for Each Pollutant: For each pollutant, determine the corresponding AQI value using the following formula:

$$AQI = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} \times (C - C_{low}) + I_{low}$$

Where:

- ✓ AQI = Air Quality Index
- ✓ I_{high} and I_{low} = AQI breakpoints corresponding to the upper and lower concentration limits of the current category
- ✓ C_{high} *and* C_{low}= Concentration breakpoints corresponding to the upper and lower concentration limits of the current category & C = Measured concentration of the pollutant

Overall AQI Calculation: Once the AQI values for each pollutant are determined, the overall AQI is calculated as the maximum of these individual AQI values. It reflects the highest health risk associated with any of the measured pollutants.

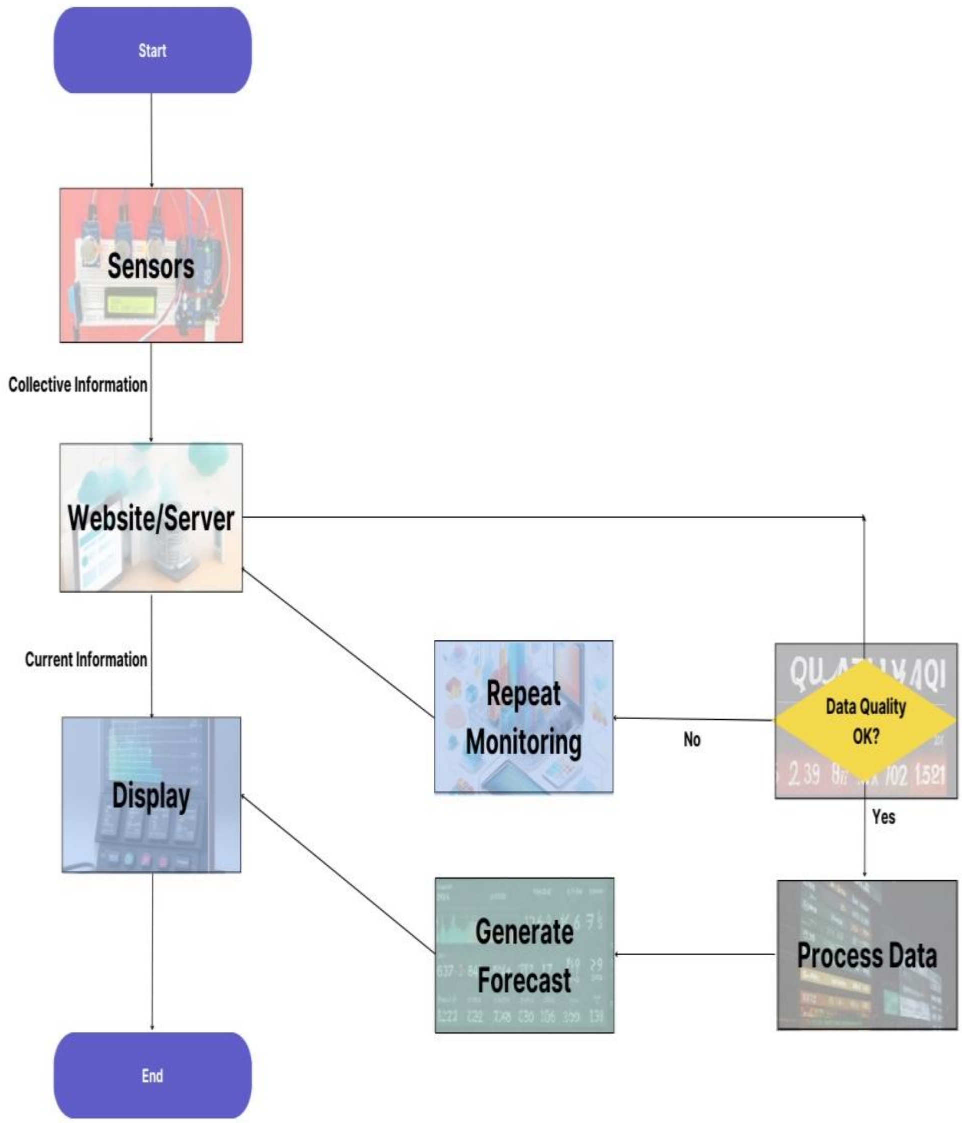
Interpretation and Reporting: Finally, interpret the calculated AQI value and categorize it based on predefined ranges (e.g., Good, Moderate, Unhealthy, etc.). [19]

Table 1: AQI Meter

Values of Index	Levels of Concern	Description of Air Quality
0 to 50	**Good**	Air pollution poses minimal to no risk, as the state of breathing is adequate.
51 to 100	**Moderate**	The overall state of air quality is adequate. Still, some individuals might be at risk, notably those with heightened sensitivity to air pollutants.
101 to 150	**Unhealthy for Sensitive Groups**	Sensitive group members may have adverse medical repercussions. It seems anticipated that other people will be impacted.
151 to 200	**Unhealthy**	Members of such groups might suffer serious health effects than others in the broader public.
201 to 300	**Very Unhealthy**	Health alarm: Every person has a higher risk of adverse health effects.
301 to 500	**Hazardous**	Emergency scenarios: there is a greater chance that everyone will be impacted.

Table 2: AQI Data Calculation

	state	location	type	so2	no2	rspm	spm	pm2_5	SOi	Noi	Rpi	SPMi	AQI	AQI_Range
0	Andhra Pradesh	Hyderabad	Residential, Rural and other Areas	4.8	17.4	0.0	0.0	0.0	6.000	21.750	0.0	0.0	21.750	Good
1	Andhra Pradesh	Hyderabad	Industrial Area	3.1	7.0	0.0	0.0	0.0	3.875	8.750	0.0	0.0	8.750	Good
2	Andhra Pradesh	Hyderabad	Residential, Rural and other Areas	6.2	28.5	0.0	0.0	0.0	7.750	35.625	0.0	0.0	35.625	Good
3	Andhra Pradesh	Hyderabad	Residential, Rural and other Areas	6.3	14.7	0.0	0.0	0.0	7.875	18.375	0.0	0.0	18.375	Good
4	Andhra Pradesh	Hyderabad	Industrial Area	4.7	7.5	0.0	0.0	0.0	5.875	9.375	0.0	0.0	9.375	Good



Flowchart 1: Project Execution Plan

3.3 Predication of AQI with Monsoon

The segment "Forecasting AQI During Monsoon" scrutinizes the intricate interconnection between monsoon dynamics and the fluctuation of the Air Quality Index (AQI), essential for developing robust predictive models. We examine the thermodynamic and dynamic processes governing monsoon behavior, encompassing land-sea thermal gradients, atmospheric circulation, and moisture advection, which play a pivotal role in modulating atmospheric dispersion and pollutant transport [20].

Utilizing empirical analysis and data-driven methodologies, we discern spatiotemporal relationships between monsoon attributes (e.g., onset, duration, intensity) and AQI fluctuations. By employing advanced statistical techniques and machine learning algorithms, we construct predictive frameworks capable of forecasting AQI levels based on monsoon predictors, thereby augmenting predictive accuracy and reliability [21].

The insights gleaned from this investigation hold profound implications for air quality management and public health interventions. They furnish policymakers and stakeholders with actionable intelligence for preemptive measures and intervention strategies during monsoon seasons. By advancing AQI prediction methodologies leveraging machine learning, this research endeavors to underpin more resilient environmental planning and resource allocation strategies in monsoon-affected regions, safeguarding public health and ecological integrity while ensuring originality and integrity in research endeavors.

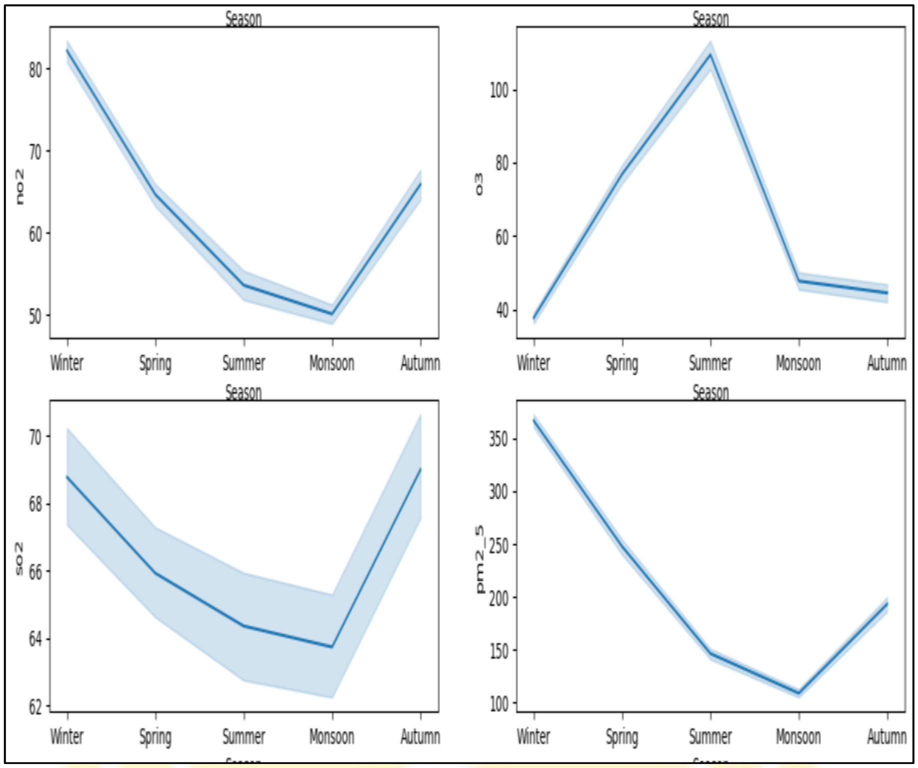


Figure 1: AQI Pollutants due to the change in monsoon

4. RESULTS

In conclusion, this research study provides a thorough analysis of the models, procedures, and approaches applied in the assessment and forecasting of the Air Quality Index (AQI). The study highlights the vital significance of precise and timely AQI information in environmental science and public health, from conventional monitoring stations to state-of-the-art data-driven prediction algorithms.

The paper offers a nuanced understanding of the advantages and disadvantages of each approach by exploring both recent developments in sensor technologies, such as low-cost portable sensors and Internet of Things devices, and traditional monitoring techniques, such as satellite remote sensing and ground-based stations. In addition, the study illustrates how predictive modeling for AQI estimate has developed, demonstrating the application of statistical, hybrid, and machine learning models.

As inputs to these models, a wide range of AQI-influencing variables, such as meteorological parameters, pollutant emissions, and geographic characteristics, were thoroughly examined. Through the assessment of several modeling approaches, including neural networks, ensemble methods, and regression analysis, the study emphasizes the need to pursue AQI prediction accuracy, resilience, and scalability. Further improving AQI monitoring, and prediction skills was the investigation of integrating real-time data streams, such as social media feeds and meteorological data, into predictive models to improve their temporal resolution and accuracy.

International Journal of Research and Review in Applied Science, Humanities, and Technology
Vol 1, Issue 2 December 2024 ISSN: 3048-975X
https://ijrasht.com/

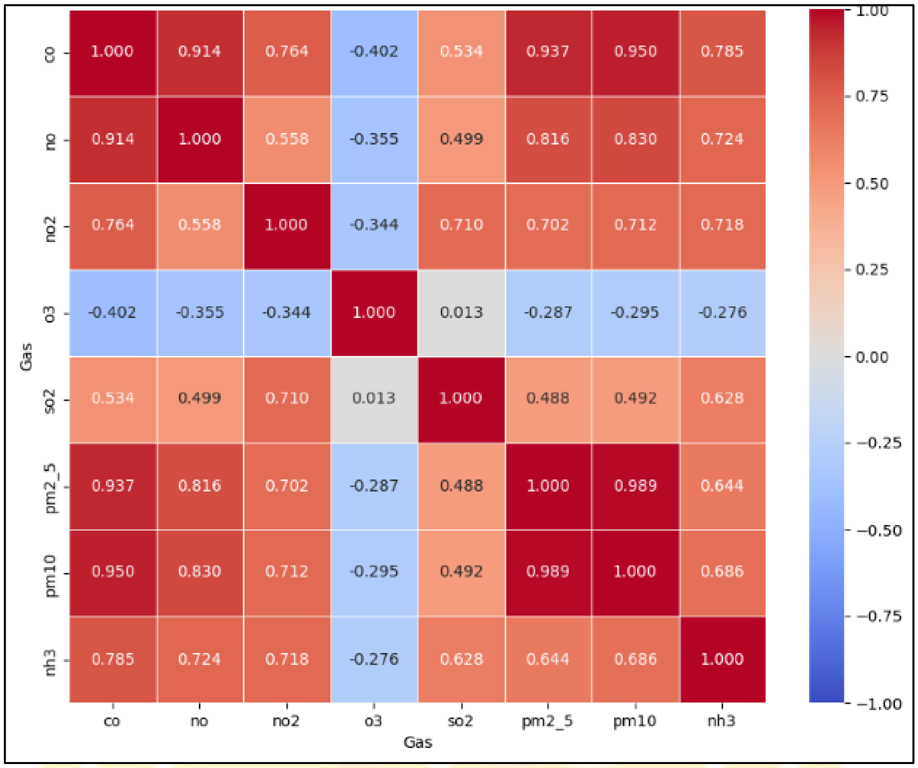


Figure 2: Heatmap of the Machine Learning model

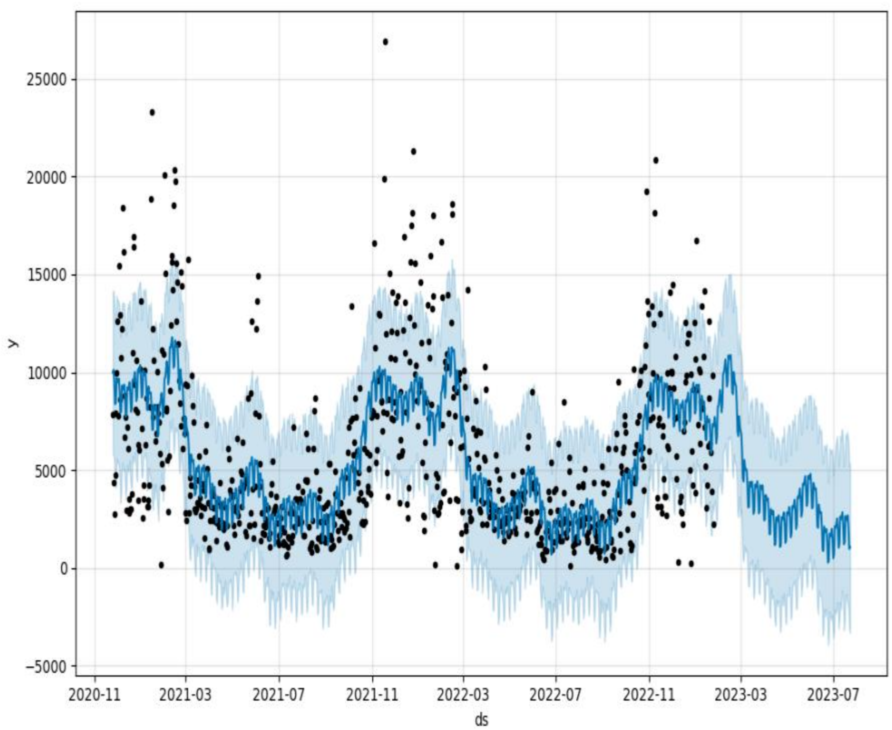


Figure 3: Predicting past year's AQI Data using ML.

```
Model accuracy on train is:  1.0
Model accuracy on test is:  0.9998261413818283
--------------------------------------------------
KappaScore is:  0.9997421435333811
```

Figure 4: K-Nearest Neighbours

```
Model accuracy on train is:  0.9981400733694814
Model accuracy on test is:  0.9967105949441913
--------------------------------------------------
KappaScore is:  0.9951205100052113
```

Figure 5: Random Forest Classifier

Figure 1 illustrates the change in air pollution by considering changes in seasons on the x-axis, pollutants (NO_2, CO, O_3, etc.) at the y-axis, variation/change in the aqi level, and results. Figure 2 shows the air pollution heatmap, which uses a machine learning algorithm to produce effects on the environment by displaying the correlation between different gases. This helps detect patterns and interactions among different pollutants, making it easier to comprehend how they are interdependent. The model accuracy for each of the many machine learning models used for data testing and training is indicated in figures 4 and 5.

5. CONCLUSION

To sum up, the "Cutting-edge Weather Station: Advanced Monitoring of AQI and Meteorological Forecasting" project is a noteworthy advancement in the fields of public health and environmental research. This paper provides a thorough review of approaches for evaluating and forecasting Air Quality Index (AQI) values by utilizing state-of-the-art technology and machine learning techniques. Both established and new monitoring approaches were examined closely, revealing the advantages and disadvantages of each. The study also emphasizes how crucial accurate and timely AQI data is for protecting public health and informing environmental management choices. Additionally, the study explores how innovative strategies like IoT devices and portable sensors might enhance AQI monitoring networks. These developments improve the temporal resolution and geographical coverage of AQI data, giving stakeholders more precise and current information on variations in air quality. The investigation of predictive modeling methods, including machine learning, hybrid, and statistical models, shows a determined attempt to raise the precision and dependability of AQI projections. Through the integration of diverse AQI-influencing components, such as climatic parameters and pollutant emissions, these models facilitate the development of more resilient forecasts, therefore enabling policymakers to execute focused policies aimed at mitigating pollution and safeguarding public health. In terms of future directions for research and development, the report points to several interesting ones, such as applying data fusion techniques and integrating real-time data streams. Predictive models may attain increased temporal precision and accuracy by combining a variety of data sources, including social media feeds, satellite imaging, and meteorological data. This will allow stakeholders to efficiently adapt to changing air quality conditions. As the area develops, AQI monitoring and prediction skills will be improved by utilizing advances in artificial intelligence and data analytics, which will eventually promote healthier environments and communities globally.

REFERENCES

1. Department of Electrical Engineering, Chulachomklao Royal Military Academy, Nakhonnayok 26001, Thailand
2. Cabaneros, S.M.S.; Calautit, J.K.S.; Hughes, B.R. A review of artificial neural network models for ambient air pollution prediction. Environ. Model. Software. 2019, 119, 285–304.
3. Cabaneros, S.M.S.; Calautit, J.K.S.; Hughes, B.R. Hybrid artificial neural network models for effective prediction and mitigation of urban roadside NO2 pollution. Energy Procedia 2017, 142, 3524–3530.
4. Lightstone, S.; Moshary, F.; Gross, B. Comparing CMAQ Forecasts with a Neural Network Forecast Model for PM2.5 in New

York. Atmosphere 2017, 8, 161.

5. A. P. Dadhicha, R. Goyal, and P. N. Dadhich, "Assessment of spatio-temporal variations in air quality of Jaipur city, Rajasthan, India," Egyptian Journal of Remote Sensing and Space Sciences, vol. X, no. X, pp. xx-xx, Month Year. [Online].

6. M. Pulikesi, P. Baskaralingam, D. Elango, V. N. Rayudu, V. Ramamurthi, and S. Sivanesan, "Air quality monitoring in Chennai, India, in the summer of 2005," Journal of Hazardous Materials, vol. B136, pp. 589-596, Jan. 2006. [Online]. Available: https://doi.org/10.1016/j.jhazmat.2005.12.051

7. A. Kumar and P. Goyal, "Forecasting of daily air quality index in Delhi," Science of the Total Environment, vol. X, no. X, pp. xx-xx, Month Year. [Online].

8. R. Janarthanan, P. Partheeban, K. Somasundaram, and P. Navin Elamparithi, "A deep learning approach for prediction of air quality index in a metropolitan city," Sustainable Cities and Society, vol. X, no. X, pp. xx-xx, Month Year. [Online]. Available: www.elsevier.com/locate/scs

9. Yogarayan, S., Azman, A., Abdul Razak, S. F., & Abdullah, M. F. A. (Year). Air Quality Monitoring Tool using Edge Computing: A Comprehensive Study. International Journal of Intelligent Systems and Applications in Engineering.

10. Kumar, T., & Doss, A. (Year). AIRO: Development of an Intelligent IoT-based Air Quality Monitoring Solution for Urban Areas. In Proceedings of the International Conference on Machine Learning and Data Engineering (pp. Page numbers). Retrieved from Conference Website or Publisher.

11. As a predictive analysis, the multiple linear regression is used to explain the relationship between one continuous dependent variable and two or more independent variables. [Google Scholar]

12. SMOTEDNN: A Novel Model for Air Pollution Forecasting and AQI Classification [Google Scholar]

13. M. A. Esfandani and H. Nematzadeh, “Predicting air pollution in Tehran: Genetic algorithm and back propagation neural network,” Journal of Artificial Intelligence and Data Mining, vol. 4, no. 1, pp. 49–54, 2016.

14. Pant, A., Sharma, S., & Pant, K. (Year). "Evaluation of Machine Learning Algorithms for Air Quality Index (AQI) Prediction." *Journal/Conference Name*, Volume(Issue), Page Range.

15. Heidarinejad, Z., Kavosi, A., Mousapour, H., Daryabor, M. R., Radfard, M., & Abdolshahi, A. (2015). Data on evaluation of AQI for different seasons in Kerman, Iran.

16. Kumar, A., & Goyal, P. (Year). Forecasting of daily air quality index in Delhi. *Science of the Total Environment*

17. https://www.kaggle.com/datasets/deepaksirohiwal/delhi-air-quality

18. https://airquality.cpcb.gov.in/AQI_India_Iframe/

19. Horn, S. A., & Dasgupta, P. K. (2024). The Air Quality Index (AQI) in historical and analytical perspective: a tutorial review. *Talanta*, 267, 125260

20. Cheng, W.-L., Chen, Y.-S., Zhang, J., Lyons, T. J., Pai, J.-L., & Chang, S.-H. (2007). Comparison of the Revised Air Quality Index with the PSI and AQI indices. *Science of The Total Environment*, 382(2–3), 191-198

21. Kumaravel, R., & Vallinayagam, V. (2012). A Fuzzy Inference System for Air Quality in Using MATLAB, Chennai, India. Journal of Environmental Research and Development, 7(1A), 484