Ecocast: AI-Driven Air Quality Forecast

Boddu Lokesh[1], Kasani Sidhartha Goud[2], Mummlaneni Chatrapathi Sai[3], Dr. Anshika Pali[4]
[1,2,3]*Computer Science Engineering (AIML) Sphoorthy Engineering College (JNTUH)* Hyderabad, India
[4]*Asst. Professor, Dept. Computer Science Engineering (AIML) Sphoorthy Engineering College (JNTUH)*

***Abstract*- This project looks at how artificial intelligence can help expect the hourly consolidation of air toxin Sulphur ozone, element matter (PM2.5), and Sulphur dioxide. As one of the most excellent procedures, AI can efficiently prepare a model on a large amount of data by using large-scale streamlining computations. Even though several works use AI to predict air quality, most of the earlier studies are limited to long-term data and easily instruct regular relapse designs (direct or nonlinear) to expect the hourly air pollution focus. This paper suggests advanced analysis to simulate the hourly environmental change focus based on the previous day's weather-related data by calculating the expectation for more than 24 hours as an execute multiple tasks learning (MTL) issue. This allows us to choose a suitable model with a variety of regularization strategies. We suggest a useful regularization that maintains the assumption patterns of concurrent hours to be near to each other, and we evaluate it to a few common MTL expected completion such as normal Frobenius standard regularization, normal atomic regularization, and 2,1-standard regularization. Our tests revealed that the suggested boundary declining concepts and constant hour-related regularizations outperform open-product relapse models and regularizations in terms of execution.**

***Keywords*— Machine learning, Air Quality**

I. INTRODUCTION (*PROBLEM STATEMENT*)

Air pollution is a major environmental concern that poses significant health risks to populations worldwide. Conventional air quality monitoring systems rely on a network of physical sensors to collect real-time data on pollutant concentrations. While these systems provide valuable insights into air quality patterns, they are often limited in their spatial coverage and may face challenges in maintaining sensor accuracy and calibration.

To address these limitations, machine learning (ML) offers promising potential for enhancing air quality monitoring and prediction capabilities. ML algorithms can be trained on large datasets of historical air quality data, meteorological data, and other relevant environmental factors to identify patterns and relationships that can be used to predict future air quality levels. Additionally, ML can be employed to analyse realtime sensor data to identify anomalies, detect potential sensor malfunctions, and improve data quality.

II. OBJECTIVES

This project aims to develop an intelligent air quality monitoring system that leverages machine learning techniques to predict and measure air quality with greater accuracy and spatiotemporal coverage. The system should incorporate the following functionalities:

1. Data Collection: The system should collect real-time air quality data from a network of physical sensors and historical data from various sources, including regulatory agencies, research institutions, and open data platforms.
2. Data Preprocessing: The system should employ data preprocessing techniques to clean, normalize, and prepare the collected data for machine learning analysis. This may involve handling missing values, removing outliers, and standardizing data formats.
3. Feature Engineering: The system should extract relevant features from the pre-processed data, including pollutant concentrations, meteorological parameters, temporal information, and spatial coordinates. These features will serve as input to the machine learning models.
4. Model Training: The system should train various machine learning models, such as regression models, time series models, and ensemble models, to predict future air quality levels. The models should be evaluated and optimized based on their performance metrics, such as accuracy, precision, and recall.
5. Air Quality Prediction: The system should utilize trained machine learning models to generate real-time and forecasted air quality predictions for

specific locations. The predictions should be presented in a clear and understandable format, such as numerical values, color-coded maps, and air quality index (AQI) values.

6. Sensor Data Analysis: The system should implement anomaly detection and sensor malfunction identification algorithms to analyse real-time sensor data. This will help maintain data quality and ensure the accuracy of air quality measurements.
7. Visualization and User Interface: The system should provide a user-friendly interface that allows users to visualize air quality data, predictions, and sensor information. The interface should be accessible through web applications or mobile applications.
8. Deployment and Maintenance: The system should be deployed in a scalable and maintainable manner. This may involve cloud-based infrastructure, containerization technologies, and continuous monitoring of system performance.

By successfully implementing these functionalities, the project will contribute to the development of advanced air quality monitoring systems that can provide more accurate, comprehensive, and timely information to support public health initiatives, environmental protection efforts, and informed decision-making.

III. MOTIVATION

The motivation behind developing a machine learning-based air quality prediction and measurement system stems from the growing concerns about air pollution and its detrimental impacts on human health and the environment. Air pollution is a major global issue, causing millions of premature deaths annually and contributing to a range of respiratory and cardiovascular diseases. Accurately predicting and measuring air quality is crucial for implementing effective mitigation strategies, protecting public health, and promoting sustainable environmental practices.

Traditional air quality monitoring methods, while essential, face limitations in providing real-time, fine-grained, and predictive insights. Machine learning offers a promising solution to these challenges by leveraging its ability to analyse large datasets, identify patterns, and make predictions. By integrating machine learning algorithms into air quality monitoring systems, researchers and environmental agencies can:

1. Real-time air quality prediction: Machine learning models can analyse historical air quality data, weather patterns, and other relevant factors to predict future air quality levels with high accuracy. This real-time prediction capability empowers individuals and authorities to make informed decisions about outdoor activities, traffic management, and emission reduction strategies.
2. Fine-grained air quality mapping: Traditional monitoring stations often provide limited spatial coverage, leaving gaps in air quality data. Machine learning techniques can infer air quality levels for areas without direct monitoring stations, generating comprehensive and detailed air quality maps. This enhanced spatial resolution enables more targeted interventions and public health advisories.
3. Identification of air pollution sources: Machine learning models can analyse complex relationships between air pollution levels and various factors, such as industrial emissions, traffic patterns, and meteorological conditions. This ability to identify pollution sources is crucial for developing targeted mitigation strategies and enforcing emission regulations.
4. Adaptive air quality monitoring: Machine learning algorithms can continuously learn from new data and adapt to changing environmental conditions, ensuring that air quality monitoring systems remain effective over time. This adaptability is essential for maintaining accurate predictions and addressing the dynamic nature of air pollution.

A. Overview

Using Machine Learning Approach, this project presents an IoT-based AQ scanning and perception framework. The harassment location stage is based on the sensor values that have been obtained. In this way, non-assaulted sensor values are tested, and the air pollution level is predicted.

B. Scope

1. Citywide Performance: The project aims to predict air quality in a city-wide level by collecting and analysing data from major pollution sources and affluence of green areas.
2. Collaboration with Environmental Agencies: Collaborating with environmental agencies to collect and analyse data generated by air quality

monitoring sensors to enhance the reliability of the model.

3. Focus on Mobility Patterns: Collecting data on traffic congestion and other factors that contribute to pollution can help to create a more accurate prediction of air quality.

C. Literature Survey

Air quality issues are a rising global concern, impacting health, environment, and economy. Thankfully, advancements in ML and IoT offer promising solutions for air quality prediction and monitoring. This survey explores the state-of-the-art in this field. Dense networks of IoT sensors collect real-time air quality data (PM2.5, PM10, CO, NO2, etc.) and environmental parameters (temperature, humidity, wind speed). This data forms the foundation for accurate prediction models. Diverse ML algorithms, including Support Vector Regression (SVR), Random Forest Regression, and Long Short-Term Memory (LSTM) networks, are employed to analyse sensor data and predict future air quality levels. Each algorithm has its advantages and limitations, making selection crucial for specific contexts.

D. Techniques and Algorithms

ML algorithms used for air quality prediction based on regression analysis:

- Random Forest (RF): Random Forest is based on the generation of several decision trees. The prediction will be the average of the predictions provided by the different trees. For the construction of each decision tree, a data sample is selected from the training dataset. The rest of the data will be used to estimate the decision tree error. The subset of independent variables that can be used for splitting each node are randomly selected. *Extremely randomized trees* (ERT) is a slightly modified random forest algorithm. Figure 1b Balogun and Tella (2022) shows a graphical representation of the general structure of a Random Forest Regressor.

- K-nearest neighbors regression (KNN): The *k*nearest neighbors algorithm is often applied to classification problems although it can be also applied to regression problems. The idea of this algorithm is simple. Given a distance (Euclidean distance, Mahalanobis distance, etc) and a k value, the algorithm calculates the distance between a data point and the training dataset points for selecting the k nearest and establishes the average of them as a prediction. An improvement of this algorithm is the algorithm known as *weighted k-nearest neighbors* (WKNN). In this case, the prediction considers a weighted arithmetic mean for calculating the prediction. Figure 1c Tella et al. (2021) shows a graphical representation of the general structure of a KNN model.

- Support vector regression (SVR). Support vector machines are mainly used in classification problems. However, they can be also applied to regression. In this case, the approach is called support vector regression. Let y and $x_1,\dots,x_p$ be the dependent variable and the independent variables, respectively. Basically, SVR works as follows. First, a linear regression function, that is, a hyperplane $h(x) = w_1x_1 +\dots+w_px_p +b$, must be defined. Then, a margin of tolerance ε is considered, expecting that all data will be at most at distance ε from the hyperplane. In case the deviation of some points is larger than this value, *slack variables* $\xi,\xi' \geq 0$ can be introduced to cope with them. The final goal is to find the minimum of the function:

$$\min \left(\frac{1}{2} \quad w \quad ^2 + C \sum_{i=1}^{N} (\xi + \xi') \right)$$

considering the restrictions : $y-w \cdot x_i - b \leq \varepsilon + \xi$ $w \cdot x_i + b - y_i \leq \varepsilon + \xi$

In case of non-linear functions, SVR uses kernel functions to transform data into a higher dimensional space in order to develop a linear regression transformation.

- Decision trees (DT). The aim of this algorithm is to design a model for predicting a quantitative variable from a set of independent variables. The algorithm is based on a recursive partitioning. Trees are composed of decision nodes and leaves. DT regression usually is built by considering the standard deviation reduction to determine how to split a node in two or more branches. The root node is the first decision node that is divided on the basis of the most relevant independent variable. Nodes are split again by considering the variable with the less sum of squared estimate of errors (SSE) as the decision node. The dataset is divided based on the values of the selected variable. The process finishes when a previously established termination criterion is satisfied. The last nodes are known as leave nodes and provide the dependent variable prediction. This value corresponds to the mean of the values associated to leaves. Figure 1a Balogun and Tella (2022) shows a graphical representation of the general structure of a standard Decision Tree.

IV. EXISTING SYSTEM

Existing air quality monitoring systems typically rely on a network of ground-based sensors that measure various pollutants, such as particulate matter (PM), ozone (O3), nitrogen dioxide (NO2), sulfur dioxide (SO2), and carbon monoxide (CO). These sensors collect real-time data, which is then transmitted to a central location for analysis and dissemination.

While existing air quality monitoring systems provide valuable information, they have several limitations. First, the cost of installing and maintaining a network of ground-based sensors can be prohibitive, especially in rural or remote areas. Second, ground-based sensors can only measure air quality at specific locations, which means that they may not provide an accurate representation of air quality across a large area. Third, groundbased sensors can be affected by local factors, such as traffic or industrial emissions, which can make it difficult to identify regional or global trends in air quality.

Drawbacks Of Existing System

1. Data quality and availability: The accuracy of ML models heavily relies on the quality and availability of training data. Inconsistent data collection methods, sensor errors, and missing values can significantly impact the performance of ML models.
2. Data imbalance: Air quality data often exhibits an imbalanced distribution, with certain pollutant concentrations being much more common than others. This imbalance can make it difficult for ML models to accurately predict rare events or extreme pollution levels.
3. Data representation: How air quality data is represented can also affect the performance of ML models. Choosing the right data representation formats and feature engineering techniques is crucial for extracting meaningful insights from the data.
4. Overfitting and underfitting: ML models are prone to overfitting, which occurs when the model learns the training data too well and fails to generalize to new data. Conversely, underfitting happens when the model is too simple to capture the complex patterns in the data, leading to poor prediction accuracy.
5. Model selection and optimization: Choosing the most appropriate ML algorithm for a particular air quality prediction task and optimizing its hyperparameters can be a challenging and timeconsuming process.

V. PROPOSED SYSTEM

Machine learning (ML) has the potential to overcome many of the limitations of existing air quality monitoring systems. ML algorithms can be trained on large amounts of historical data to identify patterns and relationships between air quality measurements and other factors, such as meteorological data, traffic patterns, and land use. This information can then be used to predict air quality levels at locations where there are no ground-based sensors.

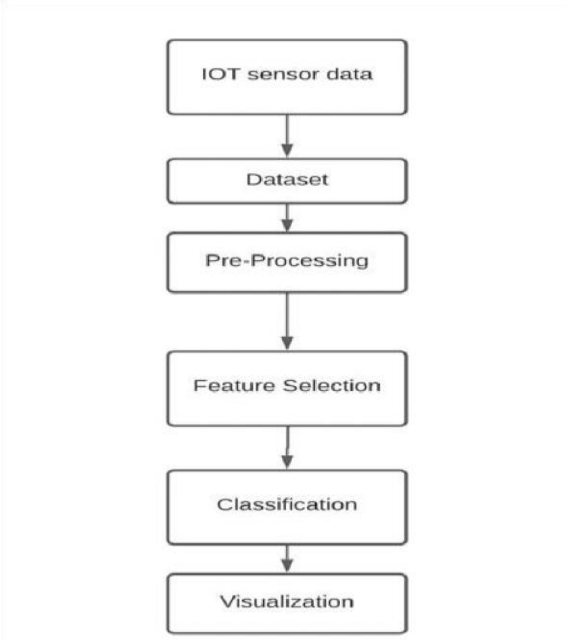
In addition, ML algorithms can be used to analyze satellite imagery and other remote sensing data to provide a more comprehensive picture of air quality across a large area. This information can be used to identify areas of high pollution and to track the movement of air pollution plumes.

Benefits of the Proposed System

The proposed air quality monitoring system using machine learning has several potential benefits over existing systems. These benefits include:

1. Increased coverage: The system can provide air quality data for areas where there are no ground based sensors.
2. Improved accuracy: The system can provide more accurate air quality predictions than existing systems.
3. Real-time monitoring: The system can provide real-time air quality data, which can be used to alert the public to health risks.
4. Long-term forecasting: The system can be used to forecast air quality levels for several days or even weeks in advance, which can be used to plan for and mitigate air pollution events.

VI. SYSTEM ARCHITECTURE



IOT Sensor Data: Internal sensors obtain data from IoT smart devices like surveillance systems, smart TVs, and wireless wearable monitors. Commercial security systems, traffic surveillance systems, and weather tracking systems all provide data. The signal is communicated, saved, and retrievable at any time. The vast majority of IoT devices produce status data gathered as original information and then used for more detailed models.

Data Set: A data class is a group of information. A data set relates with one or maybe more relational databases in the case of numerical values. Each section of the information represents a specific parameter. Each of them correlates to a particular event of the sample group in question. For each representative of the time series, the statistical model lists values with each of the different factors, such as an entity's altitude and load. Every deal is referred to as a datum. A gathering of files can also be included in a data set. Pre-Processing: In the pre-processing step, the AQI database is used. There are 'three' steps in the preprocessing process. i) Redundant data removal, ii) scriptural interpretation of missing values, and iii) data normalization. To reduce the data size, the superfluous attributes (meaningless ones) are removed in the first step. The data point of a particular component is replaced in the last thing by other constraints (CO (GT), PT08.S1 (CO), NMHC (GT), and so on). The goal of normalizing data is to give all numbers of climate the same weight.

- Delete non-essential features: assign collection (filtering and wrapper methods), object space search (see Lecture 5: Attribute-oriented analysis).

    Structural equation analysis (only for numeric attributes): finding the lowest-dimensional storage that best represents the data. Binding a smaller number of input variables.

Feature Selection: When creating a prediction model, feature extraction is the lowest quantity of effort parameters. The number of effort variables should be reduced to reduce the cognitive cost of modeling and, in certain cases, to improve the model's performance.

- MDFA: The AQI dataset contains many parameters, and inputting all of them into the classification will take a long time. As a result, the features extraction phase will be included here to choose the AQ level forecast's significant effects. To select the features of pre-processed data, the altered dragonfly optimization algorithm is used.

Air pollution is a serious issue that has resulted in thousands of premature deaths, prompting the researcher to predict air quality in terms of planning. The Air Quality Index (AQI) is a consistent place of air quality. PM10, PM2.5, NO2, SO2, CO, O3, NH3, and Pb are utilized to compute the list esteem. PM2.5 is one of them, and it is now a major source of pollution that has a significant impact on air quality. As a result, this study's main focus is on this parameter. A variety of factors influence PM2.5, so in this research, a new feature selection known as the Causality Based Practical applications was proposed to identify the most important factors influencing pollution. Classification: For attempting to solve the air quality measurement problem using e-noses, a fuzzy classification with a quantitative method is developed. The efficiency of enoses in a vibrant outdoor setting is hampered by sound, frequency drift, and a quickly shifting environment. The query is how to create predictive models that in an effective and efficient manner detect and quantitative gases. The existing research has focused on either detecting or quantifying sensing performance without taking flexible things into account. We suggest a new perfect, Indistinct Organization with Quantification Model (FCQM), in this paper to address the challenges stated above. We ran extensive tests on standard datasets produced over a currently recognized to analyse our model, and the results show that it outperforms other examine. Gas type prevention and measurement, to our understanding, is an unsolved problem.

- DLMNN Classifier: The DLMNN is given the chosen features as input here. Because the existing ANN only has one input unit (HL), training the data takes longer. To solve this problem, the proposed method employs more than three HLs. It utilizes the Gravitational Search Optimization Algorithm (GSOA), also known as DLMNN, to generate optimized
- *GSOA*: The weight values are optimized using GSOA, triggered by Newton's law of enormity and motion. The GSOA population strategies are known as agencies, and these agencies interact with one another using gravity (GF). The entities are treated as objects, and the masses evaluate their results. Because of the GF, all things start to move of other shapes with heavy groups.

For the Traveling Salesman Problem, we developed the Growing Self-Organizing Array (GSOA), which is based on personality maps' principles (TSP).

The entire protocol is a synthesis of principles used to remedy several variations of the generalized TSP with Communities (TSPN), for which the central points of the suggested unlabelled data have already demonstrated broad applicability. In terms of solution quality and computational time, the herein addressed framework ensures a functionally efficient approach that outclasses the former self map-based method for the TSP.

Visualization: After the instruction mentioned above, the semi sensor information is tested using the same training processes. Following that, the results of the evaluation classification are visualized, as explained in the segment below. The network visualizes the classified data in a variety of ways. For example, the base of prevalence in similar locations is available in a Google map with different colours, as shown below:

- Good (AQI≤50): If the order result is acceptable, the representation "Outside air is protected to inhale" is hued green.
- Moderate (AQI: 51–100): The perception "strangely touchy individuals may dodge drawn out or weighty outside effort" is featured in yellow.
- Unhealthy for the touchy gatherings (AQI: 101–150): The perception "Delicate individuals (e.g., kids, the old, outside specialists, and patients with lung sicknesses, like asthma) should restrict long or hefty open air effort" is featured in orange.
- Unhealthy (AQI: 151–200): "Delicate individuals ought to maintain a strategic distance from drawn out or substantial open air effort," says the representation. Every other person should restrict long or difficult open air action" is featured in red.
- Very unfortunate (AQI: 201–300): "Touchy individuals ought to maintain a strategic distance from all outside effort." Everyone else should restrict their open air movement" in purple. Hazardous (AQI≥301): The perception here is in maroon and says, "Everybody ought to evade all outside effort."

VII. APPLICATIONS

Environmental Monitoring:

- Air Quality Forecasting: Accurately predict future air quality levels, enabling individuals and organizations to take proactive measures to protect their health and well-being.
- Pollution Source Identification: Identify the primary sources of air pollution in a given region, allowing for targeted mitigation strategies and policy interventions.
- Real-time Air Quality Monitoring: Provide realtime air quality data, enabling people to make informed decisions about their outdoor activities and reduce their exposure to harmful pollutants.

Public Health:

- Health Risk Assessment: Assess the potential health risks associated with air pollution exposure, particularly for vulnerable populations such as children and the elderly.
- Respiratory Disease Prevention: Aid in the prevention of respiratory diseases, such as asthma and bronchitis, by providing early warnings of high air pollution levels.
- Public Health Awareness: Increase public awareness about air quality issues and promote behavioural changes to reduce exposure to air pollution.

Urban Planning and Management:

- Urban Planning and Development: Inform urban planning decisions to minimize the impact of air pollution on new developments and infrastructure projects.
- Traffic Management: Optimize traffic flow and reduce congestion to mitigate air pollution from vehicular emissions.
- Sustainable Urban Development: Promote sustainable urban development practices that

prioritize air quality improvement and public health.
Environmental Research and Policy:

- Air Quality Modeling: Develop and refine air quality models to improve predictions and understanding of air pollution patterns.
- Environmental Policy Formulation: Provide evidence-based support for environmental policy formulation and air quality regulations.
- Climate Change Adaptation: Contribute to climate change adaptation strategies by monitoring and predicting air quality changes in a changing climate.

VIII. CONCLUSION AND FUTURE ENHANCEMENT

Using the Machine Learning Approach, this project presents an IoT-based AQ scanning and perception framework. The harassment location stage is based on the sensor values that have been obtained. In this way, non-assaulted sensor values are tested, and the air pollution level is predicted. There are six different types of groups of contamination discovered and expected. The proposed strategy's results are evaluated by contrasting their speech with current methods.

In the future, the system can benefit from the integration of additional sensors to measure a broader range of air pollutants. Calibration processes can be enhanced to improve the accuracy and reliability of sensor readings. This may involve incorporating advanced sensor technologies and ensuring regular maintenance and calibration checks. To enhance the quality of data, the system could incorporate data from various sources, such as satellite imagery, meteorological data, and government monitoring stations. This would provide a more comprehensive understanding of air quality patterns and help in creating a robust dataset for machine learning algorithms.

Improving the user interface of the monitoring system can make it more accessible to a wider audience. This includes developing user-friendly dashboards, mobile applications, or web interfaces that provide real-time air quality information, historical data, and insights. Accessibility features should also be considered to ensure inclusivity. Foster collaborations with research institutions and environmental organizations to stay abreast of the latest developments in air quality monitoring. This could involve participating in research projects, sharing data for scientific analysis, and contributing to the collective knowledge of air quality management.

REFERENCE

[1] Abdullah S, Ismail M, Ahmed AN, Mansor WNW (2019) Big data analytics and artificial intelligence in air pollution studies for the prediction of particulate matter concentration. In: Proceedings of the 3rd International Conference on Telecommunications and Communication Engineering. ICTCE ’19, pp 90–94. Association for Computing Machinery, New York, NY, USA. https:// doi. org/ 10. 1145/ 33695 55. 33695 57

[2] Abirami S, Chitra P (2021) Regional air quality forecasting using spatiotemporal deep learning. J Clean Prod 283:125341. https:// doi. org/ 10. 1016/j. jclep ro. 2020. 125341

[3] Agarwal S, Sharma S, Suresh R, Rahman MH, Vranckx S, Maiheu B, Blyth L, Janssen S, Gargava P, Shukla VK, Batra S (2020) Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions. Sci Total Environ 735:139454. https:// doi. org/ 10. 1016/j. scito tenv. 2020. 139454

[4] Alhirmizy S, Qader B (2019) Multivariate time series forecasting with LSTM for madrid, spain pollution. In: 2019 International Conference on Computing and Information Science and Technology and Their Applications (ICCISTA), pp 1–5 (2019). https:// doi. org/ 10. 1109/ ICCIS TA. 2019. 88306 67

[5] Ali Shah SA, Aziz W, Ahmed Nadeem MS, Almaraashi M, Shim S-O, Habeebullah TM, Mateos C (2019) A novel Phase Space Reconstruction-(PSR-) based predictive algorithm to forecast atmospheric particulate matter concentration. Sci Program. https:// doi. org/ 10. 1155/ 2019/ 67803 79

[6] Alsaedi AS, Liyakathunisa L (2019) Spatial and temporal data analysis with deep learning for air quality prediction. In: 2019 12th International Conference on Developments in eSystems Engineering (DeSE), pp 581–587 . https:// doi. org/ 10. 1109/ DeSE. 2019. 00111

[7] Amado TM, Dela Cruz JC (2018) Development of machine learning-based predictive models for air quality monitoring and characterization. In: TENCON 2018— 2018 IEEE Region 10 Conference, pp 0668– 0672. https:// doi. org/ 10. 1109/ TENCON. 2018. 86505 18

[8] Ameer S, Shah MA, Khan A, Song H, Maple C, Islam SU, Asghar MN (2019) Comparative analysis of machine learning techniques for predicting air quality in smart cities. IEEE Access 7:128325–128338. https:// doi. org/ 10. 1109/ ACCESS. 2019. 29250 82
[9] Ao D, Cui Z, Gu D (2019) Hybrid model of air quality prediction using k-means clustering and deep neural network. In: 2019 Chinese Control Conference (CCC), pp 8416–8421. https:// doi. org/ 10. 23919/ ChiCC. 2019. 88658 61
[10] Asadollahfardi G, Madinejad M, Aria SH, Motamadi V (2016) Predicting particulate matter (PM2.5) concentrations in the air of Shahr-e Ray City, Iran, by using an Artificial Neural Network. Environmental Quality Management 25(4), 71–83. https:// doi. org/ 10. 1002/ tqem. 21464
[11] Balogun A-L, Tella A (2022) Modelling and investigating the impacts of climatic variables on ozone concentration in Malaysia using correlation analysis with random forest, decision tree regression, linear regression, and support vector regression. Chemosphere 299:134250. https:// doi. org/ 10. 1016/j. chemo sphere. 2022. 134250
[12] Balogun A-L, Tella A, Baloo L, Adebisi N (2021) A review of the inter-correlation of climate change, air pollution and urban sustainability using novel machine learning algorithms and spatial information science. Urban Climate 40:100989. https:// doi. org/ 10. 1016/j. uclim. 2021. 100989
[13] Barrón-Adame JM, Ibarra-Manzano OG, Vega-Corona A, Cortina-Januchs MG, Andina D (2012) Air pollution data classification by SOM Neural Network. In: World Automation Congress 2012, pp 1–5. https://ieeexplore.ieee.org/abstract/document /6320993
[14] Barthwal A, Acharya D (2018) An internet of things system for sensing, analysis forecasting urban air quality. In: 2018 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp 1–6 .
https:// doi. org/ 10. 1109/ CONEC CT. 2018. 84823 97
[15] Bharat Deshmukh S, Prakash Shirsat K, Prashant Dhotre S, Ravsaheb Jejurkar P (2021) A survey on machine learningbased prediction of air quality index. Int J Adv Res Innov Ideas Educ 7:1203–1207
[16] Bhatti UA, Yan Y, Zhou M, Ali S, Hussain A, Qingsong H, Yu Z, Yuan L (2021) Time series analysis and forecasting of air pollution particulate matter ($PM_2.5$): an sarima and factor analysis approach. IEEE Access 9:41019–41031 https:// doi. org/ 10. 1109/ ACCESS. 2021. 30607 44
[17] Bouakline O, ARJDAL K, KHOMSI K, SEMANE N, ELIDRISSI A, NAFIRI S, NAJMI H (2020) Prediction of daily PM10 concentration using machine learning. In: 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), pp 1–5. https:// doi. org/ 10. 1109/ ICECO CS501 24. 2020. 93143 80
[18] Bougoudis I, Demertzis K, Iliadis L (2016) HISYCOL a hybrid computational intelligence system for combined machine learning: The case of air pollution modeling in Athens. Neural Comput Appl 27(5):1191–1206. https:// doi. org/ 10. 1007/ s00521- 015- 1927-7
[19] Campalani P, Nguyen TNT, Mantovani S, Mazzini G (2011) On the automatic prediction of PM10 with insitu measurements, satellite aot retrievals and ancillary data. In: 2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp 093–098. https:// doi. org/ 10. 1109/ ISSPIT. 2011. 61515 41
[20] Chakradhar Reddy K, Nagarjuna Reddy K, Brahmaji Prasad K, Selvi Rajendran P (2021) The prediction of quality of the air using supervised learning. In: 2021 6th International Conference on Communication and Electronics Systems (ICCES), pp 1–5. https:// doi. org/ 10. 1109/ ICCES 51350. 2021. 94889 83
[21] Chen Y, Li F, Deng Z, Chen X, He J (2017) PM2.5 forecasting with hybrid LSE model-based approach. Softw Pract Exp 47(3):379–390. https:// doi. org/ 10. 1002/ spe. 2413
[22] Cheng J-C, Peng H-C (2021) Air quality forecast and evaluation based on long short-term memory network and fuzzy algorithm. In: 2021 IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII), pp 88–91. https:// doi. org/ 10. 1109/ ICKII
51822. 2021. 95747 07
[23] Chiang P-W, Horng S-J (2021) Hybrid time-series framework for daily-based PM2.5 forecasting. IEEE Access 9:104162–104176. https:// doi. org/ 10. 1109/ ACCESS. 2021. 30991 11
[24] Choi J, Kim J, Jung K (2021) Air quality prediction with 1dimensional convolution and attention on

multimodal features. In: 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), pp 196–202. https:// doi. org/ 10. 1109/ BigCo mp511 26. 2021. 00045

[25] Chowdhury A-S, Uddin MS, Tanjim MR, Noor F, Rahman RM (2020) Application of data mining techniques on air pollution of Dhaka city. In: 2020 IEEE 10th International Conference on Intelligent Systems (IS), pp 562–567. https:// doi. org/ 10. 1109/ IS483 19. 2020. 92001 25

[26] Colchado LE, Villanueva E, Ochoa-Luna J (2021) A neural network architecture with an attention-based layer for spatial prediction of fine particulate matter. In: 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA), pp 1–10. https:// doi. org/ 10. 1109/ DSAA5 3316. 2021. 95642 00

[27] Contreras L, Ferri C (2016) Wind-sensitive interpolation of urban air pollution forecasts. Procedia Comput Sci 80(C):313–323. https:// doi. org/ 10. 1016/j. procs. 2016. 05. 343

[28] Corral-Plaza D, Ortiz G, Medina-Bulo I, Boubeta-Puig J (2021) MEdit4CEP-SP: a model-driven solution to improve decision-making through user-friendly management and real-time processing of heterogeneous data streams. Knowl Based Syst 213:106682. https:// doi. org/ 10. 1016/j. knosys. 2020. 106682