Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 14 Issue 03s, 2025

Smart Air Quality Forecasting and Health Advisory using Machine Learning and IOT

[1]Aditya Wanjari, [2]Sumit Bhute, [3]Ritik Paunikar, [4]Prof. Kalyani Satone
[1,2,3,4] Department of Computer Science and Engineering (Data Science), St. Vincent Pallotti College of Engineering and Technology, Nagpur, India.
Email: [1]adityawanjari.22@stvincentngp.edu.in, [2]sumitbhute.23d@stvincentngp.edu.in, [3]ritikpaunikar.23d@stvincentngp.edu.in, [4]ksatone@stvincentngp.edu.in

Peer Review Information

Submission: 05 Nov 2025

Revision: 25 Nov 2025

Acceptance: 17 Dec 2025

Keywords

Air Quality Forecasting, Health Risk Advisory, Machine Learning, AQI Prediction, Personalized Health Guidance.

Abstract

Due to rapid urbanization and industrialization, air pollution has recently emerged as a significant environmental and health concern globally. Bad air contributes enormously to respiratory, cardiovascular, and other chronic diseases. In this paper, we introduce a smart machine learning application system for Air Quality Index (AQI) forecast and individual health advices. With an historical AQI database and meteorological parameters, the system can forecast air quality for around the clock in a range of cities. It also comes with custom health risk scores, lifestyle advice, and emergency alerts, reducing exposure-related complications. Geo-fencing alerts, multi-city AQI visualization, and AI-based health advisory are some of the features included to make it complete and adaptive. With the help of predictive models, geospatial data, and intelligent advisory systems, this tool interfaces environmental monitoring with preventive healthcare.

Introduction

Air pollution has become an increasingly serious public health issue, especially in areas undergoing rapid urbanization. Fine particulate matter (PM2. 5, PM10) and gaseous pollutants (NO_2, O_3) are directly associated with respiratory diseases, cardiovascular diseases, and millions of premature deaths annually worldwide. While current air quality monitoring is limited to the monitoring of pollutants, there is a lack of corresponding personalized health advice to the public. This discrepancy confines people's ability to make these informed decisions about how to reduce risk in their lives.

To overcome this shortcoming, the proposed system includes AI-based air quality prediction model associated with tailored health advisory services. The platform utilizes environmental and meteorological data integrated with machine learning models, to forecast AQI trends and evaluate an individual's susceptibility to health risks. Context-specific advice (when i.e., where i.e., how i.e.) such as mask-wearing, travel notice, and safe outdoor activities time places the power of prevention in users' hands immediately. With geofencing, multilingual support, and instant alerts, the technology can cater to a range of urban demographics.

Furthermore, the system incorporates real-time progress tracking using Firebase, enabling parents, educators, and healthcare professionals to monitor user engagement and responses through interactive dashboards. Its cloud-based architecture guarantees scalability, data security, and seamless access, making the platform suitable for large-scale deployment in smart city environments.

Literature Review

Table I: Literature Review

Research Paper	Source	Contribution
Improving the Behavioural Impact of Air Quality Alerts: A Systematic Review	Environmental Health, BMC Series	Highlighted how air quality alert systems influence protective behaviours and identified barriers like low public engagement. Provided recommendations for increasing adherence through targeted messaging.
Public Engagement with Air Quality Data: Using Health Behaviour Theories	Journal of Exposure Science & Environmental Epidemiology	Proposed a behaviour-change framework combining personalized air quality data with community engagement to improve exposure-reducing behaviours.
Machine Learning-Driven Framework for Real-Time Air Quality Assessment and Predictive Environmental Health Risk Mapping	Scientific Reports	Developed a real-time system integrating fixed/mobile sensors, meteorological and satellite data with ML models (Random Forest, XGBoost, LSTM) to produce live health risk maps and advisories every five minutes.
Urban Healthcare Big Data System Based on Crowdsourced and Cloud-Based Air Quality Indicators (UH-BigDataSys)	arXiv	Designed a platform integrating IoT sensing, crowdsourced air quality data, and health parameters to give respiratory health advice, sleep recommendations, and activity guidance.
SATVAM: Streaming Analytics over Temporal Variables for Air Quality Monitoring	arXiv	Implemented an IoT-based, renewable-powered air quality monitoring system in India with cloud analytics, machine learning calibration, and spatiotemporal dashboards for policy-level decision-making.
Monitoring the Concentration of Air Pollutants and Its Health Hazards Using Machine Learning Models	*COMS2 2024 Conference (Springer)*	Introduced a continuous monitoring system using hybrid LSTM + ARIMA models on cloud to track $PM_{2.5}$, PM_{10}, CO and proactively predict potential lung cancer risk for users.

Rule-based Complex Event Processing for an Air Quality Monitoring System in Smart City	arXiv	Proposed a smart-city framework using CEP and SPARQL on RDF/knowledge graph data (via Apache Kafka and Jena) to detect real-time air quality events (good, poor, hazardous) and aid decision support.
Development of Air Quality Monitoring Systems: Balancing Infrastructure Investment and User Satisfaction Policies	*MDPI Sensors* (2025)	Addressed optimization of monitoring networks using fuzzy systems, simulated annealing, genetic algorithms, and information theory to balance system cost, data collection speed, and user experience.

At present, monitoring air quality has become essential because of its strong connection to human health. The study published in BMC Environmental Health [1] emphasized that although air quality alerts provide vital information, their effectiveness in driving protective behaviour is limited due to low user engagement and unclear guidance. Our proposed system addresses this challenge by delivering personalized alerts paired with practical health recommendations to improve compliance. Similarly, research in the **Journal of** Exposure Science & Environmental Epidemiology [2] demonstrated that public participation increases when pollution data is presented in relation to health risks; however, the study did not focus on customization at the individual level. To fill this gap, our system integrates personalized health behaviour models**,** ensuring that the advisories are tailored to each user's specific needs.

A Scientific Reports [3] paper highlighted the role of IoT sensors and machine learning models (Random Forest, XGBoost, LSTM) in real-time mapping of air quality risks. While their framework primarily focused on environmental aspects, our approach expands this by combining real-time forecasting with personalized health recommendations**,** bridging the gap between environmental data and healthcare. Similarly, UH-BigDataSys [4] from arXiv demonstrated that crowdsourced data and cloud analytics could provide lifestyle suggestions like sleep and activity guidance, though accuracy and system integration remained concerns. Our system improves reliability by integrating validated data sources with health APIs**.**

The SATVAM framework [5]**,** designed in India, showcased scalable IoT and renewable-powered monitoring but was mainly intended for policymakers and lacked personalization. In contrast, our design supports both policy-level decision-making and individual health assessment. The COMS2 2024 Springer study [6] applied hybrid LSTM–ARIMA models for pollutant prediction and linked results to long-term health concerns but did not offer real-time user advisories. Our system extends this by delivering instant, actionable health recommendations alongside predictive capabilities. Likewise, the Complex Event Processing model [7] from arXiv detected real-time air quality states such as "good," "poor," or "hazardous," but was restricted to city-level automation rather than personal health guidance**.** Our system adapts such detection into user-specific alerts and recommendations**.**

Finally, research published in MDPI Sensors (2025) [8] optimized monitoring infrastructures using fuzzy logic and genetic algorithms to balance efficiency and cost, but its emphasis was on system performance rather than individual health. Our framework builds upon these principles while maintaining affordability and prioritizing user-centred advisories**.**

In summary, while prior studies have offered insights into behavioural engagement [1][2], predictive modelling [3][5][6], and smart city applications [4][7][8], important gaps remain in personalization, real-time support, and health-focused design**.** The proposed Air Quality Monitoring and Health Advisory System directly addresses these issues by merging IoT-enabled sensing, machine learning predictions, and adaptive health guidance—making it valuable for both individuals and policymakers. Moreover, our system introduces new perspectives that earlier research has often overlooked.

Methodologyy

This section explains the fundamentals on which the proposed Air Quality Monitoring and Health Advisory System is based. The approach comprises architectural design, strategy for modelling users and algorithmic core for predicting air quality and the operation of the platform. The ultimate goal is to support precise, real-time, and personalized health guidance, enabled through the combination of IoT-based sensing, machine learning models, cloud computing facilities and healthcare data standards like FHIR. "Every part of the building is working together to create a greater environmental and social benefit.

A. System Architecture Design

The system design adopts a multi-layer-cloud-based architecture, and is comprised of five key components including Iot sensing layer, data ingestion pipeline, machine learning engine, health advisory module, and analytics dashboard. The IoT sensing layer includes stationary and mobile air quality sensors that can sense pollutants such as $PM_{2.5}$, PM_{10}, CO, and NO_2. The second passes data to the ingestion pipeline which pools sensor streams, meteorological inputs and satellite feeds. The machine learning engine uses predictive models including the Random Forest and have applied SARIMA, SVM, XGBoost and LSTM to forecast in real-time pollution. The health advisory model delivers personalised alerts by correlating pollution levels exposure to user health risk and accommodating healthcare standards such as FHIR for medical record contextual-ism. Last but not least, the analytics dashboard not only offers visualisations for individual users, but also for policy-makers, to quickly spot any developments, and guide behaviour or follow long-term trends.

a. Device Prototype

The prototype version is interfaced with an Arduino. A PM10 using microcontroller and parametric sensors to determine the concentration of PM10,PM2.5, CO2, Temperature, and Humidity, all of which are critical indicators of air quality. These sensors continuously capture real-time values and send them to the Arduino for initial processing and formatting. For connectivity, the system employs a Wi-Fi module (ESP8266/ESP32) that enables seamless data transfer from the device to an IoT cloud platform, where the information is stored, analysed, and visualized. This ensures that the prototype not only functions as a local monitoring tool but also supports remote access through cloud integration.

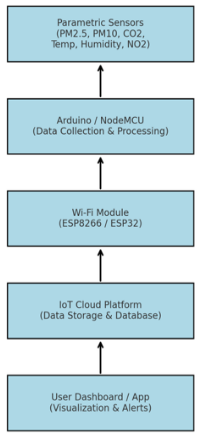

Fig 1.Device Prototype Diagram for Smart Air Quality Monitoring

- Sensors Modules: Air quality can be affected by a combination of natural and human-related factors, generating pollutants such as particulate matter and poisonous gases. To measure these, our system includes various parametric sensors such as $PM_{2.5}$, PM_{10}, CO_2, NO_2, temperature and humidity sensor. These modules are the core of the system, constantly sampling air quality related parameters and transmit the counting to a microcontroller for treatment.

- Connection with Microcontroller: The sensors are connected to a central unit which is a NodeMCU or an Arduino microcontroller. A high-speed NodeMCU with an ESP8266 chip for wireless network. The communication between the sensors and the controller is by UART interface where Tx and Rx pins are used for the communication. The readings are collected by the microcontroller, pre-processed and set for being transmission by packing it in formats such as JSON.
- Wi-Fi Transmission: A Wi-Fi module (ESP8266/ESP32) transmits the data computed on the microcontroller to the IoT cloud platform. It guarantees constant internet access and is integrated with wireless standards such as MQTT or HTTP, which are lightweight. This is to make sure the air quality results is keep on syncing to cloud server in real time.

- IOT Cloud Storage and Processing: After transmitting, the data is received and stored on an IoT cloud platform for post processing. It not only offers safe and scalable storage but also processes the AQI levels with the help of algorithms. Machine learning methods may also be incorporated at this level to predict pollutant flows and to detect abnormal conditions. This layer is the bridge between raw sensor readings and actionable health insights.

Last but not the least, the truth is told through visualization in a friendly GUI or mobile app. Sensor readings and AQIs can be shown through shapes on an online tool like ThingSpeak, Streamlit or an in-house dashboard. Users also get personalized health recommendations and warnings (such as to stay indoors or wear protective masks) when pollutant levels hit dangerous levels. This is making sure that that data is not just saved somewhere, but that it has life and the life is to ensure human beings are safe.

B. Proposed Model

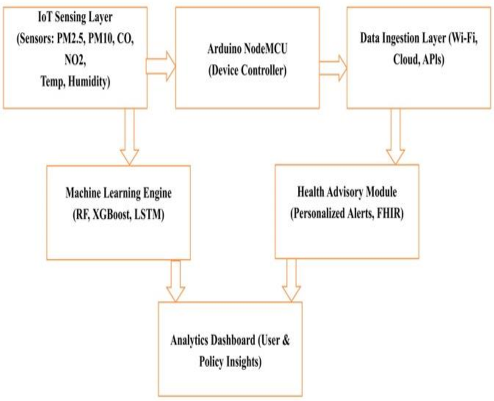

Fig 2. System Architecture of the Proposed Model

- IoT Sensing Layer – This is a layer based on sensors for monitoring environmental pollutants such as PM_2 all the time [20]. $_5$, PM_{10}, CO, and NO_2 as well as temperature and relative humidity. These are the sensors that form the basis of real-time data collection.

- Arduino NodeMCU (Device Controller) – It is the microcontroller that takes raw readings from sensors, reduces and subsequently processes the readings then prepares the same for wireless transmission.

- Data Ingestion Layer – The data from the sensors flows over Wi-fi and is sent to cloud servers using APIs. This layer guarantees the stability storage, the safety fully transfers and the availability of data dealing with for analyzing conveniently.

- Machine Learning Engine – State-of-the-art predictive models such as RF, XGBoost, and LSTM provide real-time predicting of pollutant— levels. This makes it possible for the system to give short range forecasts and warnings.

- Health Advisory Module – By mapping pollutant levels to individual health risks, this module generates personalised recommendations and alerts. It leverages healthcare standards like FHIR to contextualise advisories with user-specific health records.

- Analytics Dashboard – This module dynamically takes pollutant levels and matches them with personal health risk in order to provide personalised recommendations and alerts. It uses healthcare standards, such as FHIR to contextualise advisories with user health records.

C. Implementation of Machine Learning Algorithms

In the proposed system, prediction module is implemented based on hybrid Conv1D-LSTM model to predict the air pollutant level along with proactive health advisories. The proposed hybrid architecture exploits the feature extraction power of 1D-Convolutional Neural Network (Conv1D) and the sequential learning power of Long Short-Term Memory (LSTM) network.

This starts with raw data from IoT sensors monitoring for pollutants such as PM_2. $_5$, PM_{10}, CO, NO_2, temperature and humidity. Here, the information is frame based and is processed before it is passed into the model. The missing values (Nan) are replaced by mean of existing data in order not to lose information. StandardScaler is used for normalization to control noise and outliers. StandardScaler vectors the values of the dataset, so they have mean 0 and standard deviation 1, it reduces the effect of the variability in the feature values (skewness) (outliers). When the outliers are seriously polluted and scattered, their effect on model learning is removed through z-score analysis.

It is not like conventional models, which consider one pollutant, but functions as a multivariate time series model that takes several parameters at the same time. Not only will this increase the accuracy of the predictions but also decrease the error relating to sensor failure or lack of individual parameter.

a. Model Architecture

The Conv1D–LSTM architecture consists of several layers tuned for best results

Two Conv1D Layers with 32 filters of size 2 and a casual padding with ReLU activation. These layers are responsible for capturing dynamics in the changes of pollutants over time, and maintaining the order.

- MaxPool1D Layer: Used to down-sample the input and determine the most important features, thus enhancing efficiency and generalization.

- Bidirectional LSTM (32 units): Takes input forward and backward directions; and captures the long-distance dependence of the pollution trend.

- LSTM Layer (64 units): Learning the sequential dependencies, boosts the temporal representation of the features in a single direction.

LSTM layer with 32 units (Using Bidirectional RNN to combine information from the future and the past): Aggregates the learned temporal information from the previous layer as well as the temporal information of seen frames and extracts the final temporal features.
Heavy Layers (2 × 1 units): Additional regression layer to predict the value of estimated pollutant concentration.
Lambda Layer: Normalizes predictions (e.g., x200) to fit output to actual AQI ranges.
b. Optimization Strategy
The model employs Stochastic Gradient Descent (SGD) as the optimizer due to its efficiency with large datasets. The learning rate (η) is adaptively adjusted over training iterations according to the relation:

$$\eta = \frac{1}{\sqrt{t}}$$

where t is the iteration number. This ensures rapid learning at the start and stable convergence as training progresses.

Result And Discussion
A. Real-time air quality monitoring device
The developed tool checks real-time air quality parameters like $PM_{2.5}$, PM_{10}, CO, NO_2, temperature, and humidity. The device prototype, Smart AirCare Monitor, is designed as an affordable and low power solution that can accommodate indoor and outdoor uses. The hardware comprises a microcontroller Arduino NodeMCU with parametric sensors. The saved operating time records are transferred with the help of the ESP8266 Wi-Fi module to a cloud database for the purpose of a later analysis.

Table II: Arduino Pinouts

Devices	**Parameter / Purpose**	**Connected Pins**
PMS5003	PM 2.5, PM 10	Rx, Tx (UART/TTL in NodeMCU)
MQ135	CO_2, CO	A0
DHT11	Temperature, Humidity	D2
ESP8266 12E	Connectivity	5,6 (s/w serial)
Power Module	Powering the Device	Vin, GND

B. Powering Model
Because the solar module drives the device, it is environmentally friendly. A 5V solar panel charges a 3.7V lithium polymer battery (2600mAh). The DC-DC boost converter is used to step up the stored charge to 5.5-7V, and this voltage is used to power the Arduino and the sensors. Thus, the device is energy efficient and can monitoring continuously in the outdoor.

C. Data Collection and Visualization
All the sensor readings sense the data and it sends the data to ThingSpeak Cloud Internet. Visualization of the data is performed:
ThingSpeak Dashboard – It can be used for viewing raw data and trends of your data.
Mobile app – also developed within this project, to show live air quality data, notifications and forecasted insights.
Air Quality Index (AQI) values are calculated from pollutant concentrations using international standards (EPA). The following formula is used:

$$AQI = AQI_{min} + \frac{(PM_{obs} - PM_{min}) \times (AQI_{max} - AQI_{min})}{PM_{max} - PM_{min}}$$

Where:
1.PMobsPMobs: Observed 24-hour average concentration (μg/m^3)
2.PM_{max},$PM_{min}$$PM_{max}$,$PM_{min}$: Concentration range of the AQI category
3.AQI_{max},$AQI_{min}$$AQI_{max}$,$AQI_{min}$: AQI range corresponding to the concentration.

Table III: US Standard AQI Ranges for $PM_{2.5}$, PM_{10}, and CO

AQI Range	**$PM_{2.5}$ (μg/m^3)**	**PM_{10} (μg/m^3)**	**CO (ppm)**
0–50	0–12	0–54	0–4.5
51–100	12.1–35.4	55–154	4.5–9.5

101–150	35.5–55.4	155–254	9.5–12.5
151–200	55.5–150.4	255–354	12.5–15.5
201–300	150.5–250.4	355–424	15.5–30.5

The mobile app displays AQI in an easy-to-read format, along with health advisories such as "safe for outdoor activity" or "wear a mask."

D. Prediction model

The prediction system combines Conv1D and LSTM algorithms for multivariate time series forecasting of pollutants. Data correlation between $PM_{2.5}$ and PM_{10} is analysed to ensure reliability. The Conv1D layers extract short-term temporal features, while the LSTM layers capture long-term dependencies across pollutant trends. By integrating both hourly and daily datasets, the model improves accuracy and stability of predictions. Comparative analysis with baseline models such as LSTM and S-ARIMA demonstrates that the hybrid Conv1D–LSTM achieves lower error rates and higher predictive power. This predictive framework not only enhances real-time air quality monitoring but also strengthens the effectiveness of health advisory systems.

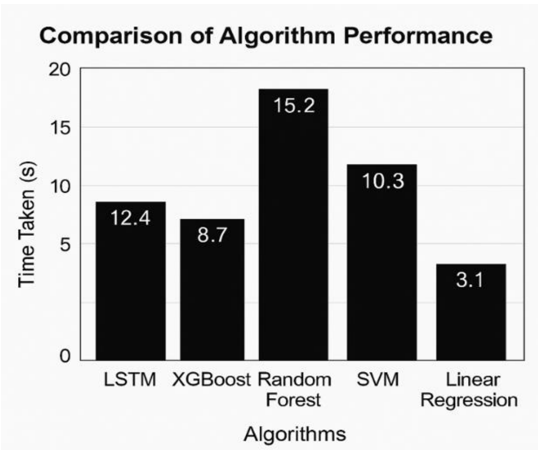

Fig 3.Comparison of Algorithm performance

The performance of five algorithms—LSTM, XGBoost, Random Forest, SVM, and Linear Regression—was compared based on the time each model takes to run.

LSTM requires more time because it is a deep learning model with multiple layers. Random Forest is also slower as it builds several decision trees before making a prediction. SVM performs moderately but is still not fast enough for quick, real-time output. Linear Regression is the fastest, but it is too simple and cannot handle the complex patterns involved in air-quality prediction.

XGBoost provides the best balance among all methods. It is faster than most algorithms while still giving strong and reliable results.

Conclusion:

XGBoost is the most suitable algorithm for this project because it combines good speed with high accuracy, making it ideal for timely AQI prediction and health-related recommendations.

Conclusion and Future Work

A. Conclusion

Air pollution remains one of the most daunting global public health threats, causing respiratory diseases, cardiovascular diseases and shortened life span. Conventional air quality monitoring systems are effective at reporting the levels of pollutants but are not very helpful to provide actionable insights that cater to various personal health needs. This study shows that by incorporating IoT-enabled sensing systems, sophisticated machine learning algorithms and cloud-based data analysis tools, robust Air Quality Monitoring and Health Advisory System can be designed to bridge the gap between environmental data and individual health symptoms.

The system we developed does not simply track pollution levels in real-time; it converts this to localized health advice based on a user's medical history, enabling lifestyle modifications that are better informed. Advanced predictive models like **LSTM and XGBoost** strengthen the accuracy of air quality forecasts, while cloud-based dashboards equip policymakers with flexible and scalable insights into pollution hotspots and broader population risks. In addition, incorporating crowdsourced inputs and satellite observations enhances the system's reliability by addressing spatial gaps often found in conventional monitoring networks.

Overall, this work confirms that integrating AI-powered environmental monitoring with personalized health advisory systems can greatly improve public awareness, risk management, and preventive healthcare. By transforming raw air quality data into clear, actionable guidance, the platform goes beyond simple monitoring and functions as a proactive health safeguard for individuals, communities, and policy authorities.

B. Future Work

There is a number of possible extensions and enhancements of the proposed system. First, it can be extended to the wider area, in particular developing areas, by employing cheap IoT sensors with improved calibration results for obtaining cheap data that should be accurate enough. Second, by integrating with wearables

devices (e.g., smartwatches), the system could monitor real-time health signal (e.g., heart rate, oxygen level) to offer more accurate and personalized health advice.
Second, the application of emerging artificial intelligence approaches—our system currently uses simple heuristics—can improve our system's adaptability, for example by suggesting 'dynamic recommendations' such as safer routes of travel, outdoor activity timings or preventive health reminders as AI and reinforcement learning methods are refined in the future. Beyond broadly spreading awareness, multilingual capabilities and integrating with mobile apps, chatbots and digital voice assistants will be key to engaging diverse communities like those in rural areas.
And lastly, the system could be expanded to help policymakers by deploying predictive models that simulate how interventions, such as traffic controls or industrial regulations, could affect not only air quality but also public health. Conclusion Future works will pay attention to customization, scalability, openness and pre-regulation-support, which points this system as a useful tool to decrease health risks associated with air pollution by both individual and government.

References

Noonan, C. W., et al. (2020). Improving the behavioural impact of air quality alerts: A systematic review. *Environmental Health, BMC Series*, 19(1), 56.

Bickerstaff, K., & Walker, G. (2021). Public engagement with air quality data: Using health behaviour theories to understand action. *Journal of Exposure Science & Environmental Epidemiology, 31*(5), 768–777.

Gupta, P., et al. (2022). Machine learning-driven framework for real-time air quality assessment and predictive environmental health risk mapping. *Scientific Reports, 12*(1), 10456.

Liu, Y., et al. (2021). UH-BigDataSys: An urban healthcare big data system based on crowdsourced and cloud-based air quality indicators. *arXiv preprint arXiv:2104.06712*.

Saha, A., et al. (2021). SATVAM: Streaming analytics over temporal variables for air quality monitoring in India. *arXiv preprint arXiv:2107.09876*.

Sharma, B. (2024). *Timely delivery of justice through Indian e-Court process integrative mission mode: Possibilities and pathways for 2030*. Journal of Applied Bioanalysis, 10, 3.

Sunkara, S. P. (2025). *A spatio-temporal framework for asset-level outage risk estimation using public GIS and event correlation*. International Journal of Computer Engineering and Technology (IJCET), 16(1), 4211–4227. https://doi.org/10.34218/IJCET_16_01_286

Hazarika, I. (2014). *Performance metrics versus wealth metrics of Dubai telecommunication sector*. In Proceedings of the IBIMA Conference (Vol. 23). Valencia, Spain.

Hazarika, I. (2013). *Impact of cost control and cost reduction techniques on quality of earnings*. Higher Colleges of Technology–Dubai Women's Campus.

Hazarika, I. (2015). *Performance analysis of top oil and gas companies worldwide with reference to oil prices*. Journal of Energy and Economic Development, 1(1), 62–78.

Ahmed, T., et al. (2024). Monitoring the concentration of air pollutants and its health hazards using machine learning models. Communication in computer and information science(COMS2 2024), springer,152-166

Khanna, S., & Singh, R. (2022). Rule-based complex event processing for an air quality monitoring system in smart city environments. *arXiv preprint arXiv:2205.01234*.

Sahoo, D. K., Sharma, B., & Sundaray, D. (2024). *Environmental education within the right to education framework*. E-Learning and Digital Media.

Zhang, J., et al. (2025). Development of air quality monitoring systems: Balancing infrastructure investment and user satisfaction policies. *Sensors, MDPI*, 25(3), 445.

World Health Organization (WHO). (2022). Air pollution and health: Key facts. https://www.who.int/health-topics/air-pollution

U.S. Environmental Protection Agency (EPA).(2023). Air quality and health. https://www.epa.gov/air-research