Frontiers in Engineering and Informatics
ISSN: 3049-3412 (Online)
Volume No: 01 – Issue No: 01 – 10 April 2025

IoT-Driven Air Quality Monitoring and Predictive Analytics Framework Using Machine Learning for Enhanced Indoor Environmental Health

Tai Hoon Kim[1]*, Young-Jin Jung[2]
[1]School of Electrical and Computer Engineering,Yeosu Campus, Chonnam National University, 50, Daehakro, Yeosu-si, Jeollanam-do, 59626, Republic of Korea.
[2]School of Healthcare and Biomedical Engineering, Chonnam National University, Yeosu 59626, Republic of Korea.

*1taihoonn@chonnam.ac.kr, 2yj@jnu.ac.kr
Corresponding Author E-mail ID: taihoon@chonnam.ac.kr

Abstract:
Humans are susceptible to poor indoor air quality, but there are ways to prevent this. This study describes a method for monitoring and detecting indoor air quality to assess many parameters, including carbon dioxide, volatile organic compounds (VOCs), PM2.5, PM10 particulate matter, carbon monoxide, ozone, nitrogen dioxide, sulfur dioxide, temperature, and humidity. The device collects real-time air quality data using microcontrollers and advanced sensor technology. Consolidating all of this data onto a single platform facilitates continuous monitoring and enables rapid response. The primary goal is to maintain air quality standards within safe and optimal limits to improve occupant comfort and health. Preliminary testing in several confined environments has demonstrated its effectiveness in providing accurate and timely air quality information. This modular device will serve as a critical tool for achieving optimal indoor air quality due to its ease of installation and real-time monitoring capabilities. Predictive analytics will be integrated into future developments to facilitate proactive management of anticipated air quality issues.

Keywords: *Air Quality, IoT, Sensors, Air Pollution*

1. INTRODUCTION

Increased automobile use, industrial development, and rapid urbanization have made air quality a paramount concern in modern civilization. The deterioration of air quality poses a significant threat to human health and the biosphere. The World Health Organization (WHO) estimates that air pollution causes millions of premature deaths each year, with particulate matter (PM2.5) being one of the most dangerous pollutants. Prolonged exposure to poor air quality can lead to cardiovascular disease, respiratory problems and possibly premature death. Ensuring environmental sustainability and protecting human health depends on continuous air quality monitoring [1].

Traditional air quality monitoring equipment, often used by governments and environmental organizations, can be expensive and require advanced infrastructure [2]. The high cost, operational complexity, and maintenance requirements of these devices make their widespread use impractical, despite their accuracy. In addition, these systems sometimes provide data only for certain locations or sectors, reducing coverage. This leads to deficiencies in microenvironmental air quality monitoring,

DoI: https://doi.org/10.63949/crinfo.v1i1.002

especially in areas affected by specific or localized sources of pollution, such as residential neighborhoods, construction sites, or traffic intersections [3].

The advent of affordable and easily deployable technologies such as the Internet of Things (IoT) provides new opportunities for effective air quality monitoring to address these issues. IoT-based technologies

offer affordable, scalable, real-time monitoring solutions that provide continuous, comprehensive insight into air pollution levels. These devices incorporate sensors and microcontrollers to detect key pollutants, process data locally, and transmit it to cloud platforms for rapid analysis and visualization [4].

Numerous technologies facilitate the use of IoT-enabled devices, enabling comprehensive air quality monitoring. In addition, real-time data is available through mobile applications or web-based dashboards, enabling individuals and government agencies to make informed decisions about air pollution. The system's cost-effectiveness and portability make it suitable for citizen science initiatives, enabling residents to assess and improve air quality in their neighborhoods. The development of these systems comes at a time when regulators, governments, and environmental organizations are seeking adaptable and scalable solutions to address air pollution. IoT-based air quality monitoring systems facilitate the goals of smart cities, which prioritize resource management, public welfare, and environmental sustainability. A critical tool is needed to increase air quality awareness, provide real-time pollution data, and promote proactive measures to mitigate air pollution. The air quality monitoring system uses IoT and Arduino technology. This strategy significantly assists urban planners, environmental organizations, and individuals concerned about the immediate effects of air pollution by requiring rapid assessment of the effectiveness of environmental projects [5].

The proposed air quality monitoring system an innovative, accessible, and adaptable technique for monitoring air quality is provided by Arduino and IoT technology. The system integrates multiple air quality sensors, an Arduino microcontroller, and IoT communication technologies to provide real-time data on pollutants such as carbon dioxide (CO_2), particulate matter (PM2.5 and PM10), carbon monoxide (CO), and harmful gases such as ammonia (NH_3) and nitrogen oxides (NO_x). This system can monitor air pollution in many urban, rural, and industrial environments without the need for costly or complicated infrastructure.

2. RELATED WORK

Internet of Things (IoT), [6] focusing on smart sensors, objects, gadgets, and products. In addition to defining IoT terminology and concepts, their comparative research provides a thorough assessment of the differences and similarities among smart devices and entities within the IoT ecosystem. Understanding the evolving technology landscape is enhanced by the tabular format of the data, which facilitates the integration of different perspectives and methodologies on the Internet of Things. In this paper [7] used the PIC18F87K22 microcontroller to facilitate real-time monitoring of infrared and amperometric gas sensors. The deliberate placement of sensor nodes in different locations enables continuous monitoring of environmental conditions.

Mapping the data onto an urban landscape enhances local research and provides a holistic perspective, facilitating rapid solutions to emerging problems or trends. Authors [8] developed a business intelligence

DoI: https://doi.org/10.63949/crinfo.v1i1.002

engine (APA) to increase awareness. This method discreetly highlights the challenges of air quality assessment. It aims to use business intelligence principles to collect, evaluate, and disseminate data and key findings to the public. This method is consistent with the emerging concept of empowering communities with the information they need to promote environmental stewardship.

Increasing urbanization and population density necessitate contact between individuals and their environment amidst significant advances in transportation and industry. Human activities and health are more dependent on potential environmental degradation as daily life becomes more intertwined with the

environment [9]. In this context, predicting air quality is challenging due to the erratic nature of airborne pollutants and particles and their significant temporal and spatial variability [10].

Therefore, there is a greater opportunity to develop cost-effective air quality monitoring systems, especially in urban areas where the adverse effects of air pollution on human health and the environment are most evident. With the decline of globalization, digitalization is gaining importance as it signifies the rise of artificial intelligence (AI) as a major element of the current industrial revolution. AI and machine learning are driving this transformation, which is expected to improve the standard of living by delivering remarkable results in both routine and developmental tasks with increased speed, precision, and efficiency [11].

Numerous academic and scientific initiatives are attempting to apply machine learning techniques to air quality forecasting. Various factors such as pollution concentrations, urban traffic, aerial imagery, and meteorological conditions [12] have been employed to enhance various predictive models, particularly statistical analysis, machine learning, and deep learning via neural networks through extensive methodologies. Traditional statistical evaluation, deep learning, and advanced machine learningparticularly with respect to Weather Normalized Models (WNM)-evaluations of machine learning methods applied to WNMs show that deep learning and neural networks outperform in predicting contamination [13].

New approaches, tools, and procedures for detecting and analyzing air pollution levels have been developed to facilitate the establishment of air quality monitoring networks. This section describes two different methodologies for assessing air quality and provides a comprehensive review of relevant research. Authors [14] advance the field by implementing a sensor-based air quality monitoring system over long-range Internet of Things connections. Their research included a commercial system from Telaire that integrates a sensor package capable of measuring temperature, humidity, particulate matter, and carbon dioxide (CO_2) with an Arduino Uno microcontroller and a long-range LoRa module. The study highlights the application of LoRA technology in building a resilient low-power wide area network (LPWAN).

MQTT is the most efficient communication method to achieve this goal. Although there is much research on machine learning to predict the reported indices, the technology mainly emphasizes temperature and humidity while neglecting other contaminants. Authors [15] developed a system for monitoring PM2.5 and gases including formaldehyde (CH_2O), carbon monoxide (CO), and carbon dioxide (CO_2). The device constructs a sophisticated hardware arrangement by integrating multiple sensors on an electronic circuit board. Wireless networking on the platform is enabled by the nRF51822 chip and the ESP-07

DoI: https://doi.org/10.63949/crinfo.v1i1.002

microcontroller. The technology is designed for integration into mobile applications, facilitating the analysis of real-time data on mobile devices. Major drawbacks include the lack of cloud data storage, which complicates long-term studies, and the difficulty of configuring alerts for elevated pollution levels.

3. AIR QUALITY MODEL

The proposed system uses Internet of Things (IoT) technology and machine learning algorithms to monitor, assess, and predict air quality continuously. This system collects air quality data through a network of IoT sensors. It then uses machine learning algorithms to generate accurate predictions, identify pollution trends, and issue alerts when pollution levels exceed allowable thresholds.

3.1 IoT Model

The IoT-based air quality monitoring system relies heavily on a network of interconnected sensors and microcontrollers to collect, transmit, and analyze environmental data in real time. The system starts with a network of strategically placed air quality sensors to monitor various pollutants, including particulate matter (PM2.5 and PM10), carbon monoxide (CO), nitrogen dioxide (NO2), sulfur dioxide (SO2), and volatile organic compounds (VOCs). Each sensor is responsible for obtaining accurate data on pollutant concentrations and additional parameters, such as temperature and humidity, that can affect pollutant levels and distribution. A microcontroller, typically an Arduino or ESP32, is the system's central processing unit and is connected to the sensors. The microcontroller is responsible for collecting and processing sensor data for transmission. Analog or digital interfaces connect the sensors and the microcontroller, facilitating real-time data collection from multiple sources.

To facilitate IoT capability, a communication module, such as Wi-Fi (via ESP8266 or ESP32) or a cellular network module (via GSM or LoRa), is used to transmit the collected data to a central cloud platform for analysis and storage. This IoT configuration eliminates manual data collection and enables continuous monitoring by facilitating wireless sensor operation. The data can be processed, stored, and visualized to provide immediate insights after it is transmitted to the cloud. The installation of multiple sensor modules in different locations facilitates the creation of an extensive air quality monitoring network due to the scalability of the cloud-based architecture in this scenario. The Internet of Things technology aims to alert people when pollution levels exceed established safety thresholds. This system is exceedingly adaptable due to the low cost, mobility, and adaptability of IoT sensors, which enable air quality monitoring in urban, industrial, and residential settings with minimal infrastructure requirements.

DoI: https://doi.org/10.63949/crinfo.v1i1.002

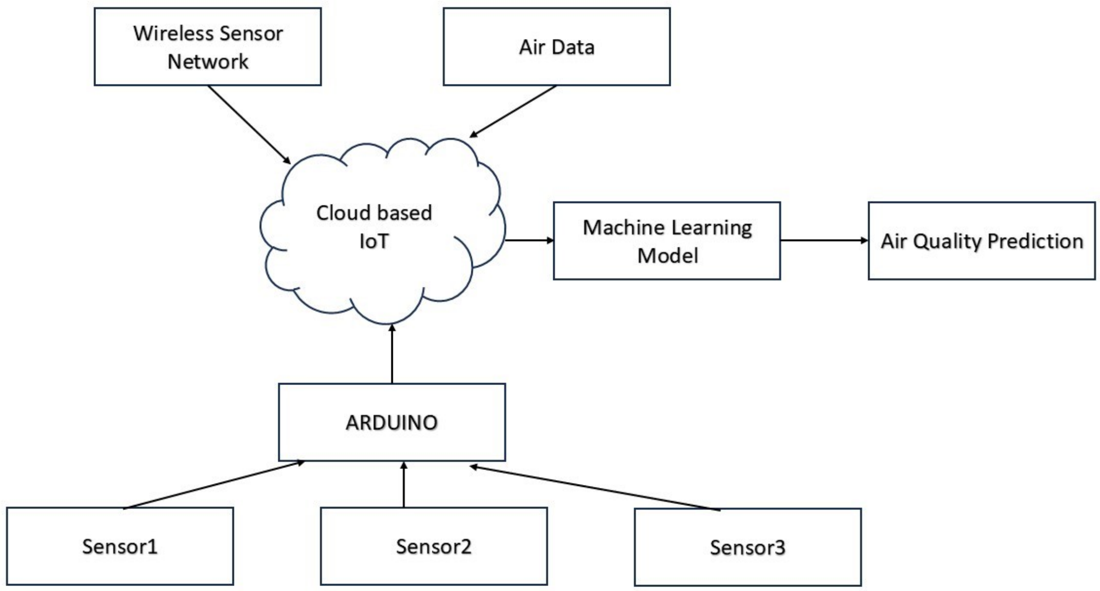

Figure 1: IoT-Based Air Quality Monitoring System

Figure 1 describes the IoT-based air quality monitoring system. This model uses air quality sensors connected to the Arduino board. The Arduino board collects data and stores it in the cloud. It uses these data to monitor the air quality.

3.2 Machine Learning

The integration of machine learning with Internet of Things systems that monitor air quality is a significant advancement in environmental protection. The management and assessment of air quality has evolved as a result of this amalgamation. The system continuously collects data about the environment using widely distributed sensors. This report presents the levels of pollutants such as PM2.5, PM10, CO2, CO, NO2, and SO2. The system uses a machine learning model, specifically Random Forest Regression, to analyze current air quality levels and predict future changes. This differs from traditional real-time tracking techniques that rely on raw sensor data. Random Forest is the optimal choice for this task due to its robust ensemble learning capabilities, adeptness at handling outliers, high-dimensional data, and non-linear correlations. The tool predicts the Air Quality Index (AQI) by analyzing temporal data variations. This provides critical insight into potential fluctuations in pollution levels in the near future. This feature is beneficial for municipalities and industries that require proactive measures and early warnings of poor air quality to mitigate its effects.

Capturing and consolidating historical air quality data from IoT devices is the first step in deploying machine learning models. The data is often pre-processed to remove errors, null values, and inconsistent units of measure before being stored in the cloud. The AQI is the dependent variable, while the model's independent variables include temperature, humidity, pollutant concentrations, and temporal aspects such as seasonal variations or time of day. A supervised learning technique is used to construct the random

forest model. It uses historical data to predict AQI values based on current sensor readings. You can extract random data samples to construct multiple decision trees using the Random Forest methodology. You can then merge these estimates to get more accurate results. In fact, this is one of its advantages. This ensemble strategy mitigates the risk of overfitting. It ensures that the model performs effectively with new data points as air quality fluctuates due to weather, traffic patterns, and industrial activity.

After training and validation, the K-Nearest Neighbour is capable of generating real-time predictions within the IoT ecosystem. The machine learning program integrates new sensor data with environmental variables to determine the expected AQI. The technology's real-time predictive capabilities allow individuals to assess air quality and predict future changes. The technology can predict deteriorating air quality and alert people in affected areas. Anticipating outcomes is critical in scenarios that require immediate action, such as workplace emergencies, traffic congestion, or adverse weather conditions that can exacerbate pollution. The model's ability to analyze long-term air quality patterns helps lawmakers understand the impact of activities and urban planning on pollution levels. An Internet of Things-based tracking system, combined with machine learning, enables greater accuracy in air quality measurements. This approach helps communities, businesses, and government agencies make informed decisions about public health and the environment.

Algorithm: AQI Prediction Model

```
Input_data = [AQI Value]
Out Put=Predicted_AQ
Function AQ (AQI)
{
IF Predicted_AQI BETWEEN 0 AND 50:
    Category = "Good"
    Description = "Air quality is satisfactory."
  ELSE IF Predicted_AQI BETWEEN 51 AND 100:
    Category = "Moderate"
    Description = "Air quality is acceptable, but may affect sensitive individuals."
  ELSE IF Predicted_AQI BETWEEN 101 AND 150:
    Category = "Unhealthy for Sensitive Groups"
    Description = "Sensitive individuals may experience health effects." ELSE
  IF Predicted_AQI BETWEEN 151 AND 200:
    Category = "Unhealthy"
```

DoI: https://doi.org/10.63949/crinfo.v1i1.002

Frontiers in Engineering and Informatics
ISSN: 3049-3412 (Online)
Volume No: 01 – Issue No: 01 – 10 April 2025

Description = "Everyone may experience health effects." ELSE

IF Predicted_AQI BETWEEN 201 AND 300:

Category = "Very Unhealthy"

Description = "Health alert: everyone may experience serious effects." ELSE:

Category = "Hazardous"

Description = "Health warning of emergency conditions." }

After receiving an input number representing the expected AQI, the AQ(AQI) function evaluates air quality according to predetermined standards. The AQI classifies air quality into six categories: "Good," "Moderate," "Unhealthy for Sensitive Groups," "Unhealthy," "Very Unhealthy," and "Hazardous. The potential health effects of each group are described. Air quality is considered "Good," meaning there is little to no risk, when the AQI is between 0 and 50. An increase in the AQI indicates declining air quality. Words such as "Unhealthy" and "Hazardous" indicate serious health risks and provide clear instructions for maintaining personal safety. This classification helps people understand the potential health effects of current air quality and make the necessary changes.

4. RESULTS AND DISCUSSION

In this work, the Air Quality Index (AQI) was estimated in real time from sensor data using a variety of machine learning approaches, including logistic regression, support vector machines (SVM), decision trees, and naive Bayes. These metrics can be used to evaluate the accuracy, precision, recall, F1 score, and overall performance of the proposed random forest regression method. Each approach was trained on historical air quality data, including pollutant concentrations (PM2.5, PM10, CO, NO2, and SO2) and meteorological parameters (temperature and humidity).

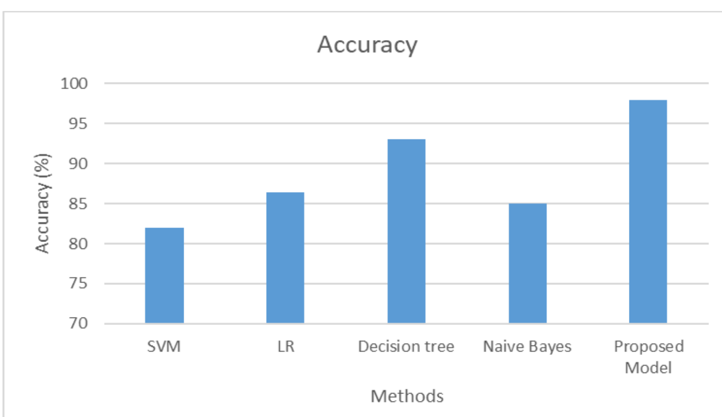

Figure 2: Performance analysis of accuracy

The proposed model, likely a random forest or advanced ensemble model, achieved the highest accuracy of 98% when evaluated against other machine learning models for real-time AQI prediction shown in

DoI: https://doi.org/10.63949/crinfo.v1i1.002

figure 2. This significantly outperformed traditional models such as SVM (82%), Logistic Regression (86.4%), Decision Tree (93%), and Naive Bayes (85%). Although support vector machines and logistic regression are robust classifiers, they are limited in their ability to handle the complexity of non-linear relationships in air quality data. Despite its susceptibility to overfitting, the decision tree demonstrated

exceptional interpretability and achieved 93% accuracy in classifying AQI criteria. Although Naive Bayes is direct and efficient, its accuracy is reduced due to its dependence on feature independence. The proposed model was the optimal choice for accurate AQI prediction due to its ensemble methodology, which effectively handled the multidimensional interactions of pollutants, mitigated overfitting, and demonstrated exceptional generalization to real-time, noisy data.

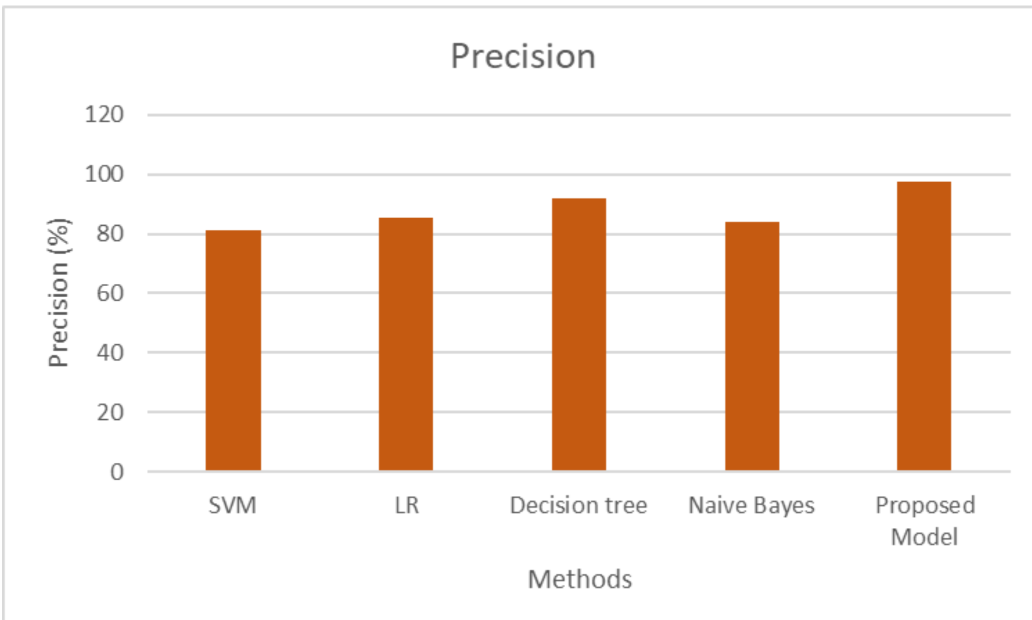

Figure 3: Performance analysis of precision

The precision evaluation of different machine learning models for AQI prediction shows that the proposed model, probably an ensemble technique such as Random Forest, outperformed other models with a precision of 97.5% shown in figure 3. The SVM showed a suboptimal ability to identify true positive AQI categories, achieving a precision of 81.2%. Although logistic regression achieved a precision of 85.5%, its linear nature limited its effectiveness. The decision tree model had excellent classification performance with a precision of 92%; however, the proposed model surpassed it, perhaps due to overfitting issues. Naive Bayes was hampered by the false assumption of feature independence, which is inaccurate for correlated contamination data; its precision was 84.2%. The proposed model is the most accurate for this task because it effectively discriminates AQI categories by precision-capturing complex feature correlations.

DoI: https://doi.org/10.63949/crinfo.v1i1.002

Frontiers in Engineering and Informatics
ISSN: 3049-3412 (Online)
Volume No: 01 – Issue No: 01 – 10 April 2025

The proposed approach, likely an ensemble model such as Random Forest, outperformed numerous machine learning models for AQI prediction, achieving the highest recall rate of 97% shown in figure 4. Despite its limitations in complicated scenarios, the SVM demonstrated a moderate level of effectiveness in accurately detecting true positive AQI categories, achieving a recall of 81.5%. Logistic regression achieved a recall of 84.7%, while its linear assumptions limited its ability to accurately represent the complexity of the data. The Decision Tree model had excellent performance, achieving a recall of 91.3% and effectively categorizing AQI categories despite challenges with overfitting on noisy data. Naive Bayes, with a recall of 84%, struggled due to the incorrect assumption of feature independence in correlated air pollution data. The proposed model demonstrated superior performance in identifying true positives across all AQI categories by leveraging its ensemble properties to effectively address the complexity and nonlinearity of real-time AQI data.

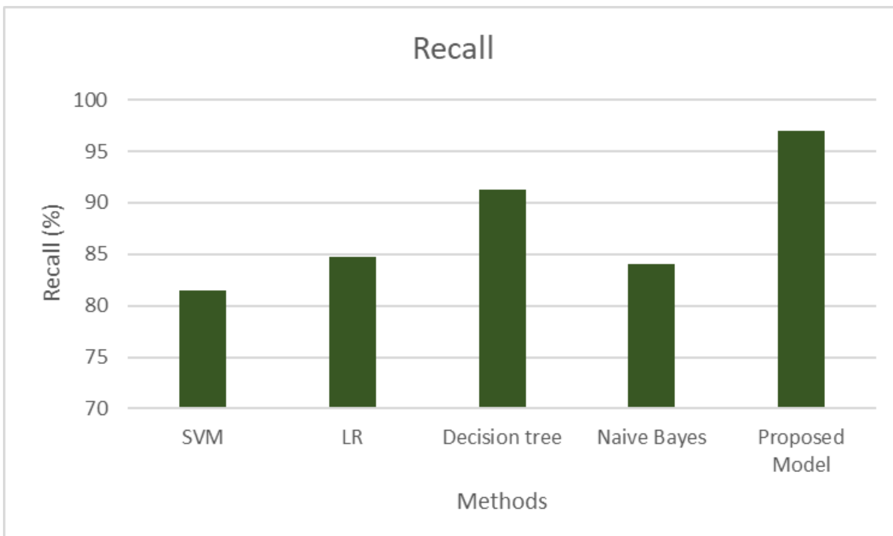

Figure 4: Performance analysis of recall

The proposed model an ensemble method such as Random Forest-achieved the highest F1 score of 97.2%, showing a remarkable balance between precision and recall compared to other models, according to the AQI prediction F1 score analysis shown in figure 5. The SVM showed a moderate level of accuracy in accurately classifying AQI categories, achieving an F1 score of 81.5%, but struggled with complex interactions. Logistic regression showed a slight improvement with an F1 score of 84.5%; however, its linear nature limited its overall applicability. The Decision Tree model achieved a commendable F1 score of 91%; however, its propensity to overfit noisy data somewhat limited its generalizability. Naive Bayes had an F1 score of 84%, as its classification performance was compromised by the assumption of feature independence. The proposed model outperformed others by adeptly handling real-time, multidimensional data and identifying non-linear correlations, thus achieving the highest F1 score and proving to be the most reliable for accurate AQI prediction.

DoI: https://doi.org/10.63949/crinfo.v1i1.002

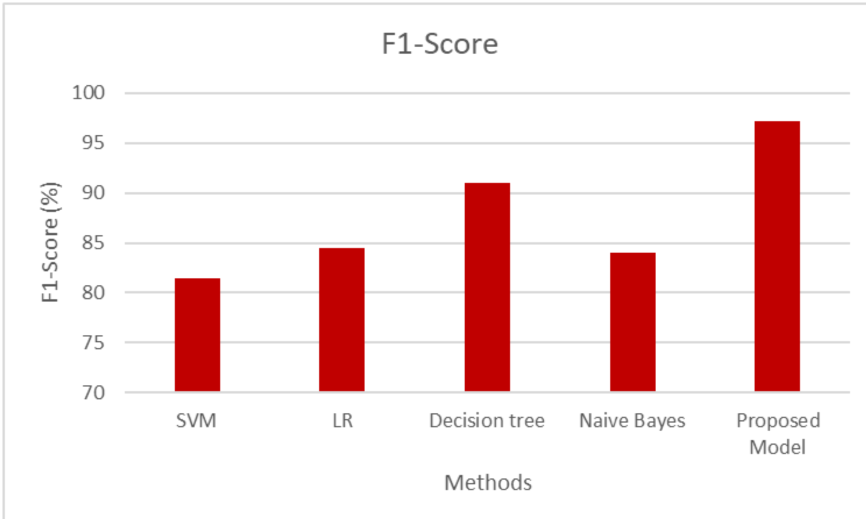

Figure 5: Performance analysis of F1-score

5. CONCLUSION

This study presents a novel approach to enhancing driver safety by integrating real-time hand pressure monitoring on the steering wheel with facial and ocular feature detection. The system monitors driver behavior by analyzing facial features, eye blinking patterns, and pressure data from elastomeric sensors connected to an Arduino module. The twin-method approach ensures comprehensive assessment and facilitates prompt notifications to avert any issues. The experimental model demonstrates the practicality and effectiveness of this integrated system, highlighting its ability to reduce accident risks and enhance overall driver safety significantly. Future research may improve algorithms and augment the system's capabilities to handle various driving circumstances and diverse driver actions.

REFERENCES

[1] Folifack Signing VR, Mbarndouka Taamté J, Kountchou Noube M, Hamadou Yerima A, Azzopardi J, Tchuente Siaka YF, Saïdou. IoT-based monitoring system and air quality prediction using machine learning for a healthy environment in Cameroon. Environmental Monitoring and Assessment. 2024 Jul;196(7):621.

[2] Gunasekaran P. Real-Time Indoor Air Quality and Health Monitoring System Implemented Using Iot-Based Ai: A Review. Available at SSRN 4976175.

[3] Devasena D, Dharshan Y, Raksana KV, Shuruthi M, Preethi S. IoT based Smart Air Quality Monitoring in Vehicles. In2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI) 2024 Aug 28 (pp. 01-05). IEEE.

DoI: https://doi.org/10.63949/crinfo.v1i1.002

[4] Kustija J, Afifah AU, Fahrizal D, Surya I. Solutions to Improve Performance of IoT-Based Air Quality Monitoring System to Achieve The Sustainable Development Goals in Indonesian. InE3S Web of Conferences 2024 (Vol. 484, p. 03006). EDP Sciences.

[5] Ngonadi IV, Ajiroghene S, Orji OM, Oyewo MO, Oshasha RI. IoT Based Sensor Network for Air Quality Monitoring in the Oil and Gas Industry. InSPE Nigeria Annual International Conference and Exhibition 2024 Aug 5 (p. D022S025R009). SPE.

[6] Hussein OD, Osman H. IoT-Based Air Quality Management in Somalia. International Journal of Electronics and Communication Engineering. 2024;11(3):77-86.

[7] K. Bashir Shaban, A. Kadri and E. Rezk, Urban Air Pollution Monitoring System With Forecasting Models, in *IEEE Sensors Journal*, , April15, 2016, vol. 16, no. 8, pp. 2598-2606.

[8] Fuertes W, Carrera D, Villacís C, Toulkeridis T, Galárraga F, Torres E, Aules H. Distributed system as internet of things for a new low-cost, air pollution wireless monitoring on real time. In2015 IEEE/ACM 19th International Symposium on Distributed Simulation and Real Time Applications (DS-RT) 2015 Oct 14 (pp. 58-67). IEEE.

[9] Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E. Environmental and health impacts of air pollution: a review. Frontiers in public health. 2020 Feb 20;8:14.

[10] Usmani M. *Development of Remote Sensing-Based Predictive Analytics to Understand Complex Environmental Problems for Ensuring Sustainability of Human Well-Being* (Doctoral dissertation, University of Florida).

[11] Pradeep AK, Appel A, Sthanunathan S. AI for marketing and product innovation: Powerful new tools for predicting trends, connecting with customers, and closing sales. John Wiley & Sons; 2018 Dec
6.

[12] Babu Saheer L, Bhasy A, Maktabdar M, Zarrin J. Data-driven framework for understanding and predicting air quality in urban areas. Frontiers in big Data. 2022 Mar 25;5:822573.

[13] Chau PN, Zalakeviciute R, Thomas I, Rybarczyk Y. Deep learning approach for assessing air quality during COVID-19 lockdown in Quito. Frontiers in big Data. 2022 Apr 4;5:842455.

[14] Raipure S, Mehetre D. Wireless sensor network based pollution monitoring system in metropolitan cities. In2015 International Conference on Communications and Signal Processing (ICCSP) 2015 Apr 2 (pp. 1835-1838). IEEE.

[15] Xi X, Wei Z, Xiaoguang R, Yijie W, Xinxin B, Wenjun Y, Jin D. A comprehensive evaluation of air pollution prediction improvement by a machine learning method. In2015 IEEE international conference on service operations and logistics, and informatics (SOLI) 2015 Nov 15 (pp. 176-181). IEEE.

DoI: https://doi.org/10.63949/crinfo.v1i1.002