Advanced Journal of Engineering and Scientific Applications (AJESA)

Research Paper Open Access

Available online at www.ajesa.com.ng Vol.2, No.3, 2025. pp.40-55

Optimized Smart City Pollution Monitoring and Control System for Awka Metropolis Using Wireless Sensor Networks and IoT

[1]Obodoeze Fidelis C., [2]Ogbonna-Mba Calister Nnenna, [3]Odegwo James I. and [4]Ikedilo Obiora Emeka, [5]Osuji Levi Nnaemeka

[1,5]Department of Computer Engineering Technology, Akanu Ibiam Federal Polytechnic,Unwana, Ebonyi State, Nigeria
[2]Department of Electrical/Electronic Engineering, Alex Ekwueme Federal University, Ndufu-Alike, Ebonyi State, Nigeria
[3]Department of Data Sciences and Analytics , Southern Delta University, Ozoro, Delta State, Nigeria

[4]Department of Computer Science Technology, Akanu Ibiam Federal Polytechnic,Unwana, Ebonyi State, Nigeria
[2]mbacally@gmail.com, [3]ifeanyijo@dsust.edu.ng, [4]ikedilo@gmail.com

Corresponding Author : **[1]drfcobodoeze@gmail.com**

Abstract

This paper presents the development and validation of a cost-effective, real-time air quality monitoring and control system for Awka Metropolis, Nigeria, integrating Wireless Sensor Networks (WSNs) and Internet of Things (IoT) technologies. By deploying pollution sensors at eight city hotspots and creating an ANN-based predictive model for the Air Quality Index (AQI), the system enables smart forecasting, real-time cloud visualization, and mobile accessibility via an Android app. Air pollution and environmental noise have become critical public health concerns in rapidly urbanizing cities like Awka, Anambra State, Nigeria. In response, this study developed and optimized a smart city air and noise pollution monitoring and prediction system using a wireless sensor network and machine learning models. A real-time monitoring substation, built with the Smart Citizen Kit (SCK), was deployed across Awka Metropolis to collect 12,958 datasets over 43 days. The system recorded six pollutant variables—PM1.0, PM2.5, PM10.0, eCO_2 , TVOC, and noise—alongside meteorological parameters including temperature, pressure, humidity, and light intensity. Eight machine learning algorithms—Multiple Linear Regression (MLR), Support Vector Regression (SVR), Multi-Layer Perceptron (MLP-ANN), Decision Tree, Random Forest, AdaBoost, XGBoost, and Extra Trees—were evaluated using both Train-Test Split and 10-Fold Cross-Validation. Results showed that ensemble models, particularly Extra Trees and Random Forest, significantly outperformed others in terms of predictive accuracy ($R^2 \geq 0.99$ for key pollutants) and residual error (low RMSE and MAE). Hybrid models (stacking and voting ensembles) did not offer additional performance gains. Noise prediction yielded poor results ($R^2 < 0.60$), suggesting the need for broader contextual features. The research contributes a deployable low-cost monitoring system and pollutant-specific predictive models that can support proactive environmental health interventions in Awka and similar urban centers.
Keywords:

KEYWORDS: *Smart city, Awka Metropolis, Air pollution, Noise pollution, Air Quality Index (AQI), Wireless Sensor Network (WSN), Internet of Things (IoT), Artificial Neural Network (ANN), Pollution monitoring, Real-time system, Smart Citizen Kit*

1. INTRODUCTION

Urbanization and industrialization, while critical to socio-economic development, have accelerated environmental degradation in many Nigerian cities. Awka Metropolis, the capital of Anambra State since 1991, has witnessed significant population growth, vehicular congestion, and expansion of commercial activities. Rapid urbanization in Awka Metropolis has accelerated vehicular emissions, industrial outputs, and population growth, culminating in elevated air and noise pollution thereby posing serious health threats including respiratory, cardiovascular, and neurological complications.

Traditional methods of air quality monitoring are often centralized, costly, and geographically limited, rendering them insufficient for real-time, localized intervention. In contrast, the advent of wireless sensor networks (WSNs), Internet of Things (IoT), and machine learning (ML) techniques presents a promising alternative. Together, they enable decentralized, cost-effective, and scalable air quality monitoring systems, capable of data-driven pollution prediction and early warning. This research paper aims to design, implement, a smart, cost-effective, real-time air quality monitoring and control system using Wireless Sensor Networks (WSNs) and Internet of Things (IoT) technologies to optimize a smart city air and noise pollution monitoring and forecasting system for Awka Metropolis. A low-cost, real-time monitoring substation was developed using Smart Citizen Kit (SCK), capturing key pollutants—PM1.0, PM2.5, PM10.0, Total Volatile Organic Compounds (TVOCs), equivalent carbon dioxide (eCO_2), and ambient noise—alongside meteorological variables. Leveraging a suite of eight machine learning algorithms, including ensemble and hybrid models, the system evaluates predictive accuracy using statistical performance metrics and cross-validation techniques.

1.1 The Study Area

After the creation of the new Anambra State from the old one in 1991, with Awka being designated the capital city, there has been increasing population, rapid industrialization and urbanization of the new capital city (Dike, 2018). This increase in population, urbanization and industrialization has led to increased human, vehicular and industrial activities that give rise to high level of environmental pollution and degradation. Awka has been besieged by lots of environmental pollution from the major environmental pollutants such as noise, particulate matters, poisonous gases such as carbon monoxide, carbon dioxide, Nitrogen oxides, Sulphur Oxides, Ozone gas, smoke etc. due to increased population, vehicular movements and industrial activities. These air and noise pollutants have been reported to cause severe human health complications and challenges. For instance, particulate matters in the air can cause health problems such as cardiovascular diseases such as cardiac arrhythmias and heart attacks, and respiratory diseases such as asthma attacks, lung cancer and bronchitis. The size of particles of these particulate matters is directly linked to their potential for causing health problems. Fine particles ($PM_{2.5}$) pose the greatest health risk (sources). Apart from these, fine particles or particulate matters can reduce visibility and cause the air to appear hazy when levels are elevated. This can cause air or road accident. Green gases such as CO_2 and O_3 (ozone) can cause severe environmental degradation and global warming. According World Health Organization (2019), about 25% of all heart diseases' deaths are caused by ambient air pollution. About 4.2 million deaths occur every year as a result of complications from air pollution diseases such as stroke, heart disease, lung cancer and chronic respiratory diseases. Fig.1 depicts the demography of number of deaths per 100,000 people in the world in 2018 arising from air pollution with Nigeria taking fourth position, ahead of China, the world's most populous country, out of twelve populated countries in the world (Forbes Statista, 2019). About 91% of the world's population lives in places where air quality levels exceed WHO limits (World Health Organization, 2019). Noise on the other hand is defined as over-loud or disturbing sound. Sound levels are measured in decibels (dB) (CPREEC, 2018). It is a unit for expressing the relative intensity of sound on a scale from 0 to 130 (Vikaspedia.in, 2020). Research has it that any sound level above 85dB can cause permanent hearing loss (Rady Children's Hospital, 2020). Noise pollution also can cause irritation and discomfort to the inhabitants where this pollution is emitted. Noise pollution can cause physiological health challenge as well as hypertension, high stress levels, tinnitus, hearing loss, sleep disturbances, and other harmful and disturbing effects. Noise pollution can come from a variety of sources such as loudspeakers, industrial machinery, aeroplanes, moving trains, construction activity or even a radio, but the three most damaging sources are industrial machinery, vehicles, and construction activity (Vikaspedia.in, 2020).

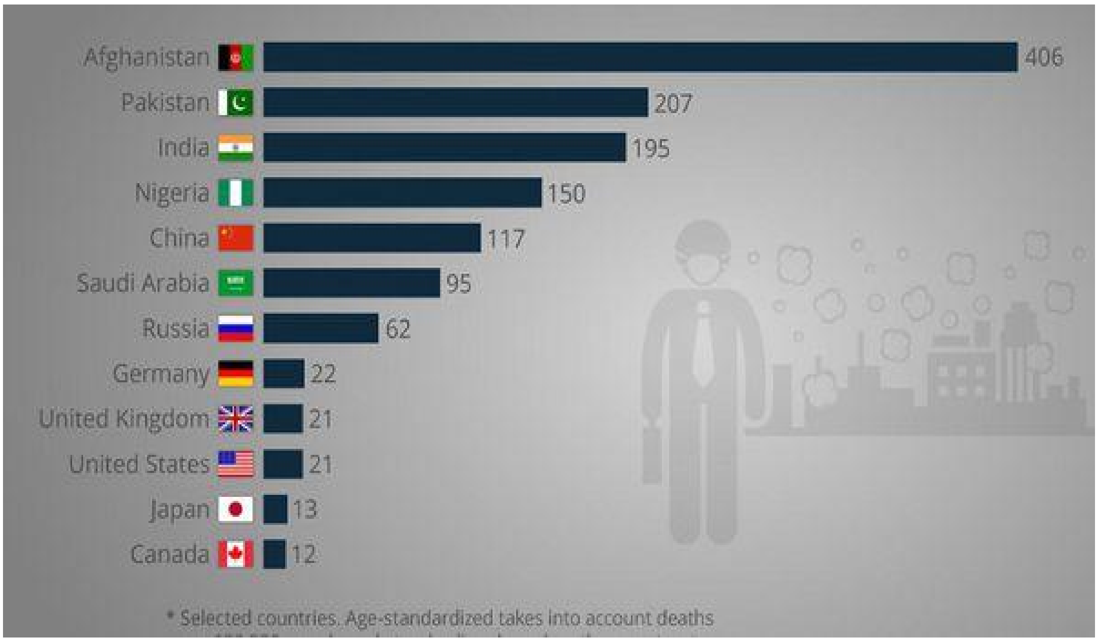

Fig.1: Statistics showing deaths from air pollution worldwide (Source: Health Effects Institute, Global Air 2018)

Several other bigger cities in the world such as Singapore, Dubai, Milton Keynes, Southampton, Amsterdam, Barcelona, Madrid, Stockholm, Copenhagen, China and New York have had this type of environmental challenge of air pollution but were able to overcome them with the aid of smart city air quality or air pollution monitoring and control system (I-Scoop; Siemens, 2019). Fig.2 depicts the current area view of Awka city from Google Map (Wikimedia, 2016).

Smart city concept (i.e. the use of intelligent or smart solutions to optimize infrastructure and engage the citizens in the management of their city) can be applied in the area of air quality or air pollution management of a city or metropolis to manage or control the extent of air pollution by giving the opportunity to residents, transportation planners, environmental regulators, and the general public to bring about proactive preventive measures for safety. The concept of smart city can be different in terms levels of implementation from city to city or country to country. This implementation depends solely on the level of development, willingness to change, reform, resources and aspirations of the city residents and the governments.

Fig. 2: Area view of Awka Metropolis from Google Map (Source: Wikimedia, 2016)

Smart city is dependent on technology but it is not defined by that technology (Obodoeze and Amadi, 2018). A smart city is an information hub that empowers citizens, supports businesses, and inspires community innovation (Saratoga Smart City Springs, 2019). Smart city concept assists the government to monitor, track and manage the people, the environment and infrastructure via connected smart objects such as sensors, Close Circuit Television (CCTV) cameras, Radio Frequency Identification (RFID) receivers, Global Positioning System (GPS) devices, etc. such that the data collected from these smart objects will be used by government to make vital and timely decision that will make life better for its citizens. Smart city concept encompasses the integration of technology, people and government policies.

1.2 The Need for Smart Air Pollution Monitoring

Urbanization and industrialization have led to increased emissions of harmful pollutants like PM2.5, PM10.0, CO, NO_2, SO_2, and O_3, which pose serious health risks. Traditional monitoring stations, while accurate, are expensive and limited in spatial coverage. This creates a need for low-cost, scalable, and real-time systems that can provide granular data across a city.

1.3 Role of Wireless Sensor Networks (WSNs)

WSNs consist of spatially distributed sensor nodes that collect environmental data and transmit it wirelessly to a central server or cloud platform. In the context of air quality:

- **Sensors** detect pollutants and environmental parameters (e.g., temperature, humidity).
- **Microcontrollers** process the data locally.
- **Communication modules** (e.g., ZigBee, LoRa, Wi-Fi) transmit the data.
- **Gateways** aggregate and forward data to cloud servers for storage and analysis.

WSNs are favored for their **low power consumption, modularity, and ease of deployment**, making them ideal for dense urban environments.

1.4 System Architecture Overview

A typical smart air quality monitoring system includes:

- **Sensor nodes**: Deployed across the city to collect data.
- **Edge processing**: Basic filtering and preprocessing at the node level.
- **Cloud platform**: For storage, visualization, and advanced analytics.
- **ML engine**: Trained on historical data to predict future pollution levels.
- **User interface**: Web dashboards and mobile apps for public access and alerts.

1.5 Real-World Applications

- **Ibarra City, Ecuador**: A WSN-based system classified urban zones into high, medium, and low pollution using ML, achieving 95% classification accuracy.
- **India**: IoT-based systems have been deployed in urban areas to monitor PM2.5 and NO2 using ML to forecast pollution and support public health planning.

1.6 Contributions of the research

This research presents the design, optimization, and implementation of a smart city air pollution monitoring and control system tailored to Awka Metropolis, Nigeria, utilizing Wireless Sensor Networks (WSNs) and Internet of Things (IoT) technologies. It addresses rising air and noise pollution caused by rapid urbanization, vehicular traffic, and industrial activities in Awka. The system integrates low-cost sensors to detect pollutants such as PM2.5, PM10.0, PM1.0 particulate matters as well as TVOC, eCO, and weather parameters such as temperature, humidity, and noise. The data is transmitted wirelessly via ZigBee and Wi-Fi modules to a cloud-based IoT platform, where it is analyzed and visualized in real time. An Android app was also developed to provide public access to real-time air quality information, while alerts are sent via SMS if pollutant thresholds are exceeded. A predictive model using Artificial Neural Networks (ANN) was developed and trained using historical air quality data. The ANN model forecasts Air Quality Index (AQI) for Awka and was validated using simulation tools like Python. The system was tested across eight pollution hotspots in the city and compared with existing systems in terms of cost, efficiency, data accuracy, modularity, and energy efficiency. A solar-powered backup ensures uninterrupted data sampling.

2. RELATED WORKS

Air quality monitoring systems from developed regions such as CAIRNet (New Zealand) and SensorWebBike (Italy) demonstrate promising frameworks. However, limitations like high cost, low modularity, and lack of localized predictive capabilities make them less adaptable to cities like Awka. Prior research (Brienza et al., 2015; Penza et al., 2017) outlined the importance of customizing sensor nodes and communication protocols for localized environmental factors.

Karth (2019) developed an IoT based Air Pollution monitoring system using Arduino Uno open source board and ATmega328 microcontroller. The system deployed MQ-135 gas sensor to detect and measure Ammonia (NH_3), Nitrogen Oxide (NO_X), Alcohols, Benzene (C_6H_6), smoke and carbon dioxide (CO_2). The ESP8266 Wi-Fi module was used to establish wireless connection between the devices. The system was connected to open source cloud server, ThingSpeak, to store and visualize the sensors' data on real time. The system worked well but has several shortcomings – only one sensor, MQ-135, was used to monitor and detect several gases at the same time. The MQ-135 sensor cannot measure each of the pollutant gas distinctly and clearly without some supporting gas sensors. Particulate matters were also not considered by this system. Also, there is no effort to backup the supply using solar energy in order to support the battery energy of system.

Yamunathangam et al (2019) proposed and implemented an IoT Enabled Air Pollution monitoring and Awareness Creation system. The system used Arduino Uno board and ATMega328 microcontroller integrated with Gas sensors – MQ-135, MQ-7, MQ-2, particulate matter sensor DSM501A and a humidity sensor to detect and measure hazardous gases such as Ammonia, Carbon Monoxide (CO), Methane (CH_4), smoke, etc. as well as particulate matter ($PM_{2.5}$) floating in the air. An Ethernet shield was connected to the Arduino Uno microcontroller to connect the sensors' data to an open source cloud server, ThingSpeak, for storage and visualization by Android users. The results of the experiment seem okay as visualized via Android App and Arduino Ethernet shield. The system did not detect and measure other poisonous gases such as SO_2, NO_2, Ozone (O_3) and other particulate matters such as $PM_{1.0}$ and PM_{10}. There is no power backup such as solar to support the battery power used by the system as power supply.

Vardhan et al (2019) proposed a Smart Industrial Pollution Monitoring system using IoT. The system used Arduino Uno based open source board and ATMega328 microcontroller to interface about five sensors- MQ-7, M213, LM35, SY-HS220 and LDR to measure and detect Carbon Monoxide (CO), Noise, temperature, Relative humidity and light intensity respectively of an ecological environment. The Arduino Uno board and microcontroller was connected to a Wi-Fi module, ESP8266, to send the sensors' data to an internet web page. The system integrated several pollutant sensors but still left out the major industrial pollutant gases such as CO_2, NO_2, O_3, SO_2 and particulate matters. The system did not mention how the sensors' data was deployed to a cloud server. The security of the sensors' data was not also considered in this system.

Choodarathnakara et al (2019) proposed and developed a real-time IoT based Air and Noise Pollution Monitoring system to measure and detect air and noise pollution in an area and alert relevant authorities when there is a breach of the standard thresholds. The system employed Arduino Uno board and ATMega3280 microcontroller for data processing and to interface about five sensors- MQ-135 gas sensor, KY038 noise sensor and DHT11 temperature sensor. The software sub-system of the system was developed using Arduino IDE V1.85, NETBEANS IDE and MySQL DBMS to test the sensors' data. The experiment posted good results in the test environment and Google Map. However, the authors have not tested the system on a live cloud server in real-time. Equally, the MQ-135 gas sensor is not adequate to cover the entire spectrum of pollutant poisonous gases such as CO, SO_2, NO_2, O_3, etc.

Aktaruzzaman et al (2019) proposed and developed an Arduino Uno Air quality measurement meter with Digital Dashboard using smartphone and Blynk IoT server to measure and detect current levels of air pollutants such as smoke, benzene, NO_2, steam, CO_2 and other harmful gases. It used MQ-135 gas sensor to accomplish that as well as SDS011 sensor to measure and detect $PM_{2.5}$/PM_{10} particulate matters. DHT11 was also used to measure temperature. All the sensors were connected to Arduino Uno board and ATMega328P microcontroller to process the sensors' data. Arduino Uno Ethernet shield hardware as used to connect the sensors' data to an enterprise-based Blynk Cloud server to store and visualize the sensors' data in real time. The system posted good results on Blynk Cloud server. The drawbacks are that MQ-135 sensor alone cannot measure and distinguish the different level of concentrations of the pollutant gases distinctly; extra gas sensors are needed to solve the problem. Equally, the system operates only on public power supply; there is no battery backup power and no solar backup. Public power supply may not be reliable for such city-wide air pollution monitoring campaign.

Mujawar and Deshmukh (2019) developed and implemented a smart environment monitoring system using both wired and wireless sensor networks (WSN) separately to measure and monitor environmental parameters such as temperature, humidity, light intensity and air quality in real-time. The authors demonstrated the two separate systems (wired and wireless) and proved that the wireless system is better in terms of cost, ease of installation, better connectivity (less or no wiring issues) and higher efficiency. This system consisted of a portable and efficient environment monitoring system based on LabVIEW GUI. The LabVIEW (Laboratory Virtual Instrument Engineering Workbench) and an Arduino IDE were used to program the system. The wireless system was designed using ATMega 328 microcontroller integrated with XBee S2 protocol for form sensing phenomena. A radio transceiver mechanism was used as a communication part accompanied by the sensors and a battery to provide power supply unit in this system. Four sensors (LM 35, SH-HS-220, MQ-135 and LDR) were used to measure temperature, humidity, air quality and light intensity respectively within Solapur University campus in India. The developed WSN system used two motes and one sink node. The system yielded good results when viewed on computer but the web reports or data from the monitoring system cannot be viewed on a mobile phone or Android device using LABVIEW software. Secondly the system does not have an energy scavenging power source to provide uninterrupted power in the case of a city-wide monitoring campaign. There is need here to develop a mobile or Android App to view the web reports.

BlessyEvangelin and Thurai (2019) proposed an IoT based air pollution monitoring system to create a smart environment. The system used an Arduino UNO and microcontroller based board, MQ-7 Gas sensor, DTH11 Humidity/temperature sensor and an LCD display to report or display the readings from the sensors. MQ-7 gas sensor can detect Carbon Monoxide, Methane and LPG very easily and quickly. The system created air quality checking and perception framework precisely estimated the grouping of poisons carbon monoxide, carbon dioxide, smoke and residue in climate but it lacks basic connectivity to cloud server and hence cannot allow users on computers and clients such as Android Apps or mobile devices to connect and view the sensor readings. Secondly, the system does not cover most of the pollutants that affect urban city. There is also no security for the protection of the data on the cloud.

Zode et al (2019) proposed and developed an Arduino based Smart Air Pollution Detection System to measure and detect only carbon dioxide gas in three cities of India-Nagpur, Madhy Pradesh and Delhi. The MQ-7 CO sensor was used to sample CO concentrations in these cities. Arduino Uno open source board and ATMega328 microcontroller were used to interface the CO sensor. The sensor data was displayed on 16x2 LCD display and a Wi-Fi module, ESP8266, was used to send the sensors' data to an open source ThingSpeak IoT cloud server. The system seems simple and low cost but it did not cover adequate number of air pollutant gases such as NO_2, SO_2, O_3, smoke, CO_2 and particulate matters ($PM_{1.0}$, $PM_{2.5}$, PM_{10}).; covering only CO is not okay for a city-wide air pollution or air quality measurement.

Gómez et al (2019) proposed and developed a Novel Air Quality Monitoring Unit Using Cloudino and FIWARE Technologies. The Cloudino is a cloud-based IoT server. The system is a smart city application designed to measure air quality using Cloudino, Arduino and Fiware platforms. The system employed the use of several air pollutants and environmental sensors such as CCS811 Sensor (to measure VOCs and CO_2 (Carbon Dioxide)), MQ-7 Sensor was used for detecting CO concentrations in the air, MQ-135 Sensor was used for detecting many gases such as Ammonia (NH_3), Nitrogen Oxides (NO_X), alcohols, aromatic compounds, Hydrogen Sulfide, Benzene, Steam, etc. while GA1A1S202WP Sensor was used as an analog light sensor to save electricity, BME280 Sensor (allows to measure the barometric pressure, altitude, temperature, and relative humidity).The system seems good in terms of simplicity, cost and the number of pollutants gases it measured and monitored but it did not cover particulate matters at all. Moreover, there was no power backup in terms of solar energy to support the project.

Menezes et al (2019) proposed and designed an IoT based Air and Sound Pollution Monitoring System to detect and measure air and noise pollution in an area. Two air pollution sensors were used in this system check the presence of harmful and hazardous gases/compounds. Such harmful gases and compounds detected in this system include Methane, propane, Butane, alcohol, noxious gases, and carbon monoxide. These harmful gases were sensed in this system using the following air pollutant sensors - MQ-135 (used to detect and measure Methane, propane, Butane, alcohol, Smoke, steam, noxious gases) and MQ-7 used to detect Carbon monoxide (CO). A DHT11 temperature and humidity sensor was used to measure temperature and relative humidity of the ambient environment. The system equally monitored sound pollution using a MIC sound sensor. The system runs on Arduino Uno based open source board with ATMega328 microcontroller and connected to ESP8266 Wi-Fi module to interface the sensors' data to an open source NodeMCU cloud server IoT platform in real time. The system was deployed both indoor and outdoor-offices, buildings etc. A decision-tree flowchart-like algorithm was used in the system as visual and analytical decision support tool to maximize the efficiency of the decision-making of the system. The system seeks okay for a versatile, real-time, flexible and cost efficient tool to monitor both air and sound quality in a particular area. The only drawback is that particulate matters were not considered in the air quality monitoring.

Ahasan et al (2018) proposed, developed and implemented an Arduino-based Real Time Air Quality and Pollution Monitoring System. The system employed two major sensing chips MQ-135 and LM393 to detect and measure several gaseous pollutants such as NH_3, NO_2, benzene, alcohol, smoke and CO_2. Arduino Uno board, a modern micro-controller board based on the ATmega328P micro-controller was used for the data processing and hosting of the sensors. Different measurements of the prototype were taken by the authors in different locations and different readings were obtained on pollutant concentrations. The system seems okay for its intended objectives but lacks the platform, IoT or cloud service to monitor the sensors' data in real-time

3. MATERIALS AND METHODS

The research paper takes a comparative experimental approach to evaluate the performance of various machine learning models for air and noise pollution prediction in Awka Metropolis, Anambra State, Nigeria.

3.0 The hardware infrastructure and Cloud server ecosystem

The Smart Citizen Kit (SCK) and the Cloud server were used to collect and measure the air pollutant variables from different hot-spots within Awka Metropolis spanning 43 days and obtaining about 12958 rows of datasets from the air pollutant environment sensors. The datasets were uploaded in real time every 5 seconds to the Smart Citizen Server, which comprises the online server with processing capability and back-end database server. The SCK is designed in a **modular** way, with a central *data logger* (the Data Board) with Wi-Fi connectivity, a micro SD-card slot, a micro USB, and a battery connector. Various sensors can be connected, either via the Urban Board, featuring different onboard sensors and a particulate matter sensor connector, or through the Auxiliary port (Smartcitizen, 2025). It is an open-source, modular device designed for **environmental monitoring**—perfect for researchers, educators, and citizen scientists alike.

Figs 3, 4, 5, 6 and 7 show the Smart Citizen Kit components and platform used in this experimental research work.

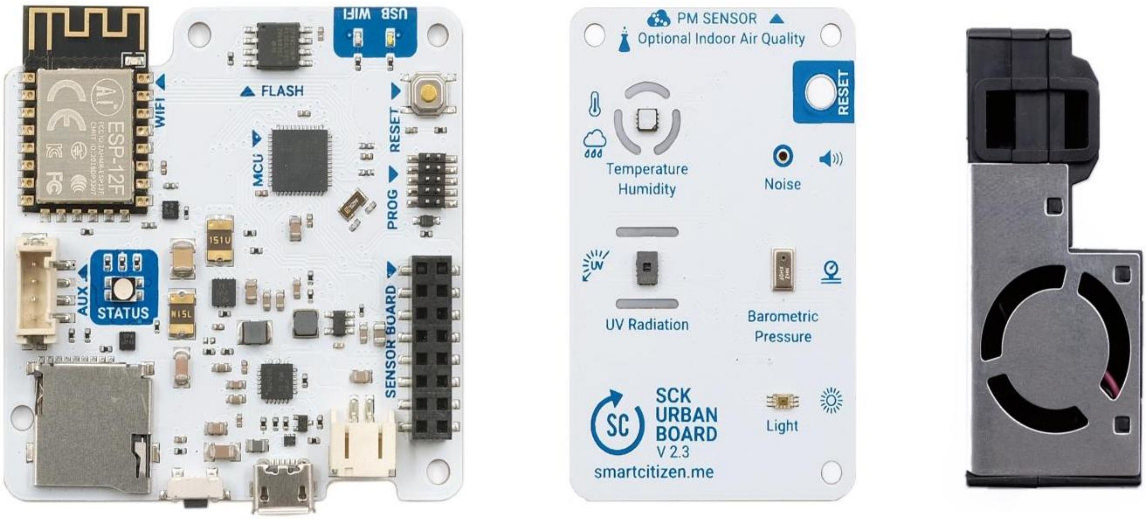

Fig. 3: The Smart Citizen Kit (SCK) main board and Urban Sensor Board and PM sensor board (Source: Smartcitizen, 2025)

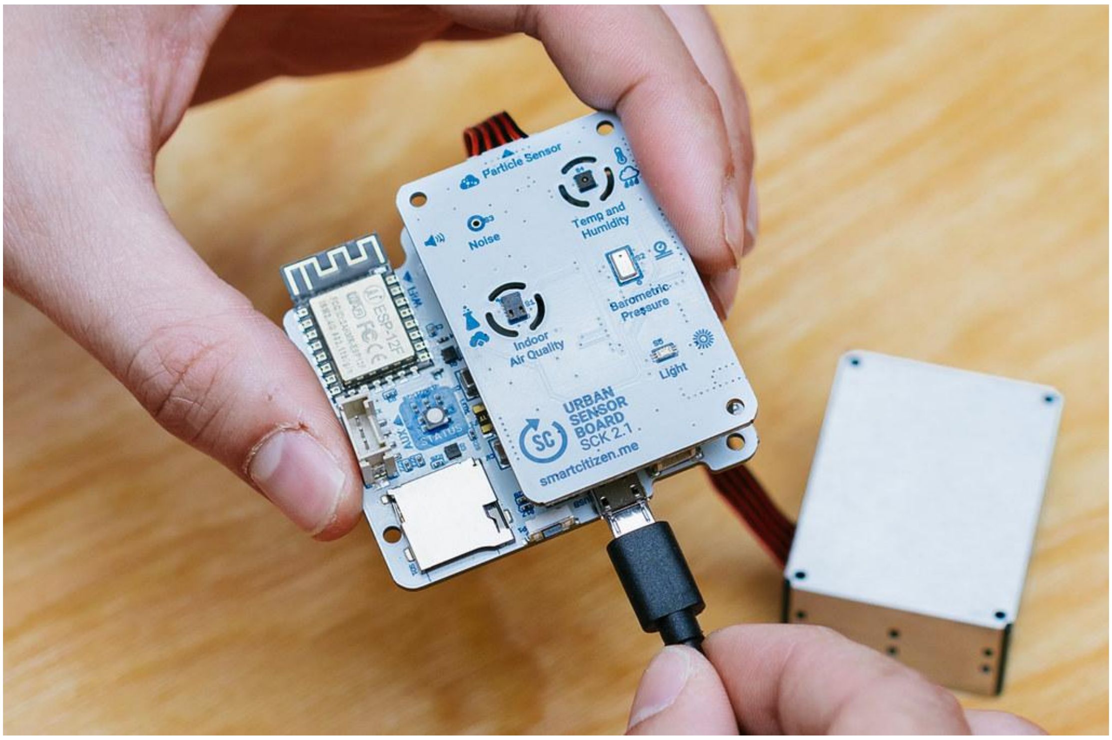

Fig. 4: The Smart Citizen Kit (SCK) Assembly (Source: Smartcitizen, 2025)

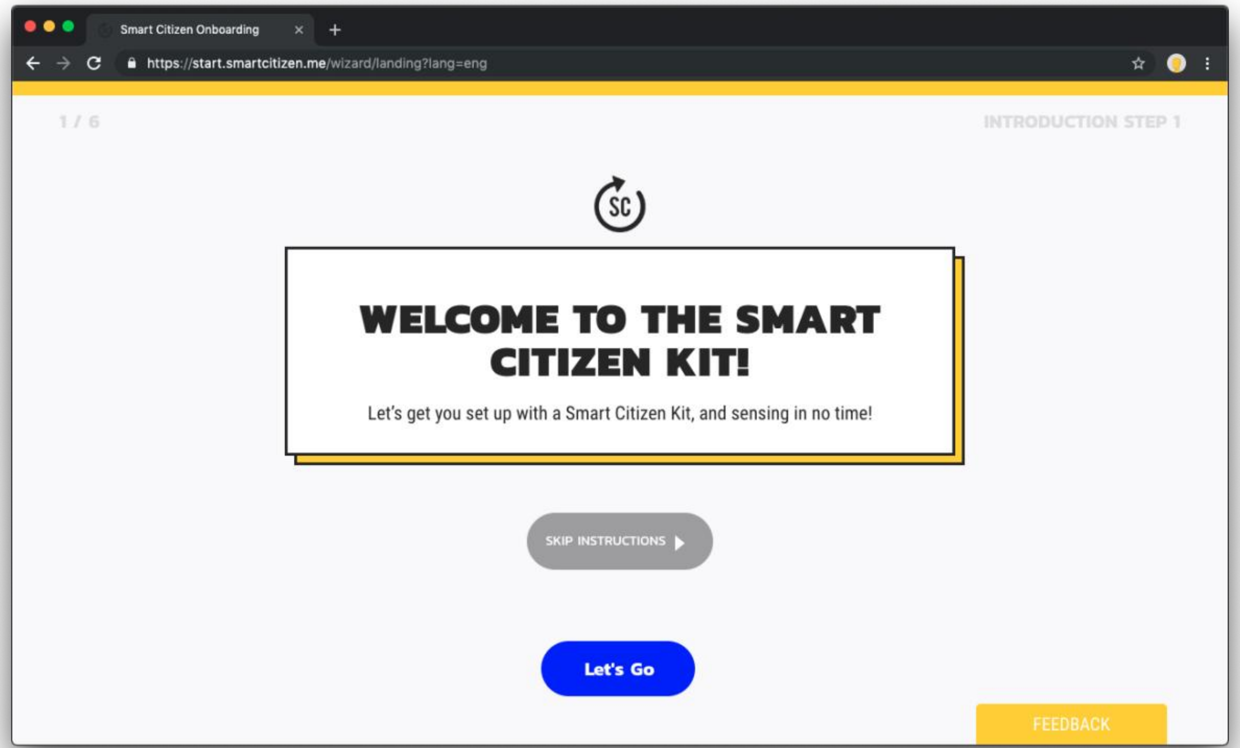

Fig. 5: The Smart Citizen Kit (SCK) successful on-boarding into the Cloud server platform

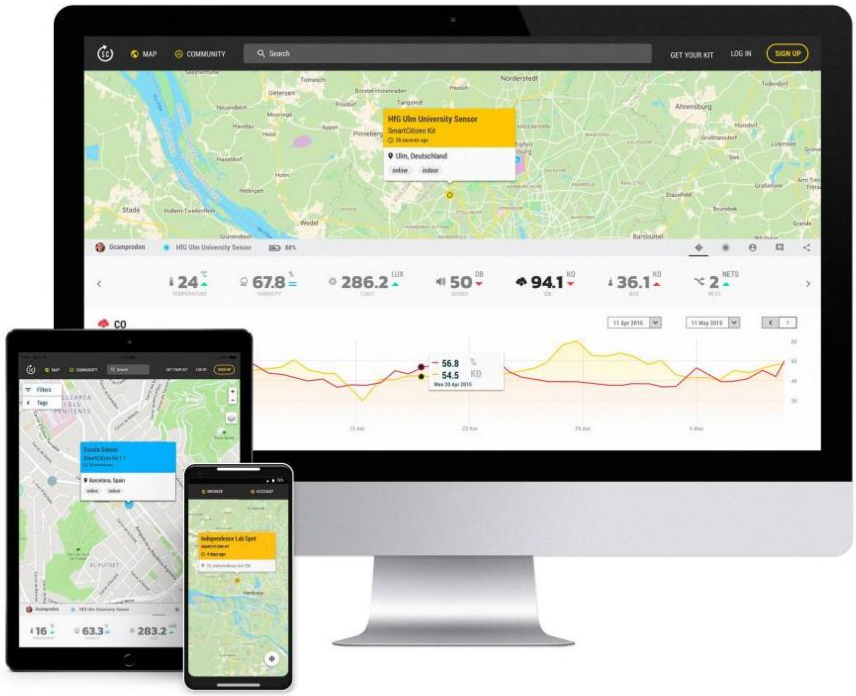

Fig.6: The Smart Citizen Kit (SCK) Cloud server platform with sensor readings displayed in desktop computer, mobile phone and tablet (Source: Smartcitizen, 2025)

Fig. 7: The deployed outdoor Air pollution monitoring system using the Smart Citizen Kit infrastructure (Source: Smartcitizen, 2025)

Published bimonthly by the Nigerian Institute of Electrical and Electronics Engineers (NIEEE) Awka Chapter, Nigeria..

The SCK comprises the following components:

A. **Data Board**: The brain of the kit, featuring:

- MicrocontrollerS
- Wi-Fi connectivity
- MicroSD card slot
- USB and battery connectors
- **Urban Board**: Houses onboard sensors and connectors for:
  - *Particulate Matter (PM1.0, PM2.5, PM10)*
  - *Noise levels*
  - *Ambient light*
  - *Temperature & Humidity*
  - *Barometric pressure*

B. Connectivity & Data

- Sends real-time data to the Smart Citizen Cloud server Platform
- Data is visualized through a web interface and accessible via open API
- Can operate **online (Wi-Fi)** or **offline (SD card mode)**

C. Power & Portability

- Rechargeable battery with ~12 hours of autonomy
- Can be powered continuously via USB for long-term deployments

3.1 Data Collection via field deployment

Field Deployment: Physical test-beds were installed at eight hot-spots across Awka Metropolis to measure and collect real-time data from the atmosphere outdoor of the air pollutant variables such as (PM1.0, PM2.5, PM10, eCO2, TVOC and noise and data analysis and modeling were conducted using these sensor data to perform machine learning predictive modeling experiments in order to develop and validate the developed ML models using the simulation output. Some samples of TVOC and eCO2 were collected indoors in some locations.

3.2 Hybrid Methodology Design

The methodology smartly combines *real-world sensor deployment* with *data-driven machine learning modeling*. It builds a robust foundation by first deploying a custom-built, low-cost **Smart Citizen Kit (SCK)** air and noise monitoring station and then collecting a 43-day dataset from different hotspots within Awka Metropolis.

3.3 Experimental Structure

A total of **ten primary experiments** and numerous sub-experiments were executed, targeting specific pollutants (PM1.0, PM2.5, PM10, eCO_2, TVOC, and noise) and using both *Train-Test Split* and *10-Fold Cross-Validation* methods. This dual-evaluation strategy boosts the reliability of the findings.

3.4 The Machine Learning Model Diversity

Eight machine learning models were compared—including **Multiple Linear Regression (MLR)**, **MLP-ANN**, **SVR**, **Decision Tree**, and four ensemble models (**Random Forest, AdaBoost, XGBoost, Extra Trees**). Two hybrid models (stacking and voting ensembles) were also tested.

3.5 The Experimental tools and test-bed

The research utilized Python 3.6.3 libraries such as **Scikit-learn, Keras, Tensorflow, and Pandas** within **Jupyter Notebooks** and Anaconda environments. The implementation choices align well with modern best practices in machine learning workflows.

3.6 Machine learning models used

Traditional ML models: These include the following: Multiple Linear Regression (MLR), Decision Tree, SVR and MLP-ANN.

Ensemble models: These include Random Forest, AdaBoost, Extra Trees and XGBoost.

Hybrid models: These include Stacking and Voting Ensemble models.

3.6 Evaluation Metrics

Three major metrics were used for performance evaluation. They include the following:

- R^2 Score (accuracy of prediction)
- MAE (Mean Absolute Error) and
- RMSE (Root Mean Square Error)

3.7 Validation and Evaluation techniques

Techniques: 70:30 Train-Test Split was used to split the datasets into training and test/validation respectively. 10-Fold Cross Validation was used to perform validation test for the developed ML prediction models.

4. RESULTS AND DISCUSSION

This section presents the results and results analysis.

Table 4.1: PM2.5 Prediction Results

Model	R^2 Score	RMSE (µg/m³)	MAE (µg/m³)
Random Forest	0.9886	0.9207	0.2802
Extra Trees	0.9886	0.9175	0.2847
XGBoost	0.9870	0.9814	0.3133
AdaBoost	0.9854	1.0386	0.3076
MLR	0.9856	1.0319	0.4470
MLP ANN	0.9815	1.1711	0.6018
Decision Tree	0.9782	1.2709	0.3812
SVR	0.9622	1.6745	0.5218

Table 4.2: PM10.0 Prediction Results

Model	R^2 Score	RMSE (µg/m³)	MAE (µg/m³)
MLR	0.9834	1.2413	0.4223
MLP ANN	0.9828	1.2637	0.5885
Extra Trees	0.9800	1.3652	0.4305
Random Forest	0.9803	1.3531	0.4313
XGBoost	0.9779	1.4319	0.4664
AdaBoost	0.9731	1.5803	0.4619
Decision Tree	0.9634	1.8457	0.5486
SVR	0.9582	1.9719	0.5876

Table 4.3: PM1.0 Prediction Results

Model	R^2 Score	RMSE (µg/m³)	MAE (µg/m³)
Random Forest	0.9713	0.8952	0.2996
Extra Trees	0.9704	0.9087	0.3084
XGBoost	0.9688	0.9326	0.3318
AdaBoost	0.9664	0.9677	0.3068
Decision Tree	0.9538	1.1350	0.3629
MLP ANN	0.8900	1.7500	0.8900
SVR	0.8139	2.2782	0.9386
MLR	0.8037	2.3396	1.6277

Table 4.4: TVOC Prediction Results

Model	R^2 Score	RMSE (ppb)	MAE (ppb)
Extra Trees	0.9994	1.3288	0.2781
XGBoost	0.9986	2.0058	0.4257
Random Forest	0.9951	3.8083	0.3157
AdaBoost	0.9924	4.7380	0.3061
Decision Tree	0.9924	4.7380	0.3061
MLP ANN	0.8747	19.2120	8.7866
MLR	0.8604	20.2781	10.0145
SVR	0.7713	25.8615	4.7753

Table 4.5: eCO_2 Prediction Results

Model	R^2 Score	RMSE (ppm)	MAE (ppm)
Extra Trees	0.9992	7.9876	2.3454
Random Forest	0.9992	7.9876	2.3454
XGBoost	0.9990	8.8319	3.0228
AdaBoost	0.9991	0.1763	1.8840
Decision Tree	0.9988	9.4314	2.3917
MLR	0.8377	111.6920	78.5114
MLP ANN	0.3700	219.2600	158.2300
SVR	0.0462	270.7918	171.1014

Table 4.6: Noise Prediction Results

Model	R^2 Score	RMSE (dB)	MAE (dB)
Random Forest	0.5859	4.3146	3.2334
Extra Trees	0.5597	4.4491	3.3584
XGBoost	0.5557	4.4694	3.3909
AdaBoost	0.5289	4.6023	3.3989
Decision Tree	0.2080	5.9670	4.3401
MLP ANN	0.2160	5.9367	4.8178
SVR	0.2508	5.8034	4.6714
MLR	0.1590	6.1489	5.0261

4.1 Results Discussion

The research conducted rigorous testing, and the outcomes are both insightful and contextually relevant.

4.1.1 PM2.5 Prediction

Random Forest and **Extra Trees** tied for best performance with $R^2 \approx 0.9886$, showing high accuracy. MAE and RMSE scores were impressively low, making these models reliable for forecasting PM2.5.

4.1.2 PM10 and PM1.0

Interestingly, **MLR** had a competitive edge for PM10, suggesting a strong linear relationship in that dataset. For PM1.0, **Random Forest** again led with excellent predictive performance.

4.1.3 eCO2 and TVOC

Extra Trees came out on top, outperforming even hybrid models. Predictions were near-perfect ($R^2 > 0.99$), likely due to clean sensor data and solid signal patterns in the collected dataset.

4.1.4 Noise Prediction:

This was a weak spot. All models—including ensemble and hybrid approaches—struggled, yielding low R^2 scores (< 0.6). This points to either insufficient input features or a need for domain-specific noise predictors like traffic volume, distance from roads, or time-of-day factors.

4.1.5 Hybrid Models (Voting & Stacking)

Surprisingly, hybrid models did not outperform the best-performing single ensemble models. This is a valuable insight, reinforcing that in some cases, complexity doesn't necessarily lead to better accuracy. It's a good reminder that model performance is data-dependent.

4.2 Summary of experiments and Results

The section summarizes the results in this research work. The concise summary of the results from your research on smart city air and noise pollution prediction in Awka Metropolis:

4.2.1 The Data and Scope

Dataset: 12,958 samples collected over 43 days
Monitored Variables:

- **Air pollutants**: PM1.0, PM2.5, PM10.0, TVOC, eCO_2
- **Weather**: Temperature, Humidity, Pressure, Light
- **Noise levels**

Tools: Python (Scikit-learn, TensorFlow, Keras), Smart Citizen Kit (SCK)

4.2.2 The Overall Best Models

Extra Trees and **Random Forest** consistently delivered the highest prediction accuracy and lowest residual errors (MAE & RMSE) across multiple pollutants.

Multiple Linear Regression (MLR) outperformed others for PM10 due to a strong linear relationship in the data.

Hybrid Models – Not Always Better: Voting and Stacking ensembles underperformed compared to single ensemble models. Complexity doesn't guarantee better performance — simpler, well-matched models excelled with clean, structured data.

Noise Prediction Challenge: All models struggled to accurately predict noise levels ($R^2 < 0.6$). This suggests a need for additional input features like traffic density, timestamps, spatial location, and land use etc. to make noise pollution prediction better and more accurate.

Table 4.7 summarize the pollutant specific model recommended for use in Awka Metropolis for monitoring different air pollutant concentrations.

Table 4.7: Pollutant-Specific Model Recommendations

Pollutant	Best Model	Notes
PM2.5	Random Forest / Extra Trees	High precision ($R^2 = 0.9886$)
PM10.0	MLR	Best fit due to linearity
PM1.0	Random Forest	Top accuracy with low errors
TVOC	Extra Trees	Nearly perfect match ($R^2 = 0.9994$)
eCO_2	Extra Trees	Strongest consistency and precision
Noise	None	Needs new features and sensors

5. CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This research successfully developed and optimized a smart air pollution monitoring and control system tailored to the specific needs of Awka Metropolis. The combination of WSN, IoT, ANN-based forecasting, and mobile interfaces makes it adaptable for other rapidly urbanizing cities with similar environmental challenges. The system provides a scalable, low-cost smart city solution for pollution monitoring. Integration of WSN, IoT, mobile access, and Machine Learning (ML) predictive analytics tailored to Awka enables a practical pathway toward cleaner and safer urban environments in Nigeria and beyond. This research is methodologically solid and reflects a meticulous, engineering-minded approach to smart city analytics. The experimental rigor, use of real-world data, and thoughtful comparison of algorithms make the findings not only academically valid but also practical for real-world deployment.

5.2 Recommendations

The following recommendations are suggested for the future applications of this work towards effective deployment and usage in Nigeria

1. Wider deployment across southeastern Nigeria
2. Integration with environmental policy and urban planning databases
3. Future work to include federated AI and edge computing for decentralized processing

References

Aarthi, A., et al. (2020). Air Quality Prediction Through Regression Model. International Journal of Scientific and Technology Research, 9(3), 923–928.

Aceves-Fernández, M.A., et al. (2020). Evaluation of Key Parameters Using Deep Convolutional Neural Networks for Airborne Pollution (PM10) Prediction. Hindawi.

Aljanabi, M., et al. (2020). Ground-level Ozone Prediction Using Machine Learning Techniques: A Case Study in Amman, Jordan. International Journal of Automation and Computing.

Allam, Z. (2019). Achieving Neuroplasticity in Artificial Neural Networks through Smart Cities. Smart Cities, 2, 118–134.

Anambra State–Nigeria Erosion and Watershed Management Project (ANS-NEWMAP). (2014). Environmental and Social Management Plan Draft Report for Nerose Plaza–Thomas Aquinas Gully Erosion Intervention Site.

Arroyo, P., Herrero, J. L., Suárez, J. I., & Lozano, J. (2019). Wireless Sensor Network Combined with Cloud Computing for Air Quality Monitoring. Sensors, 19(3), 691.

Brienza, S., Galli, A., Anastasi, G., & Bruschi, P. (2015). A low-cost sensing system for cooperative air quality monitoring in urban areas. Sensors, 15(6), 12242–12259.

Castelli, M., et al. (2020). A Machine Learning Approach to Predict Air Quality in California. Hindawi Complexity.

Chiou-Jye, H., & Ping-Huan, K. (2018). A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities. Sensors, 18(2220).

Cortina-Januchs, M.G., et al. (2015). Development of a Model for Forecasting of PM10 Concentrations in Salamanca Mexico. Atmospheric Pollution Research, 6, 626–634.

Published bimonthly by the Nigerian Institute of Electrical and Electronics Engineers (NIEEE) Awka Chapter, Nigeria..

Dedovic, M.N., et al. (2016). Forecasting PM10 Concentrations Using Neural Networks and Systems for Improving Air Quality. XI International Symposium on Telecom (BIHTEL).

Devi, P.D. & Chandrakala, M. (2021). Air Quality Prediction Using Ensemble Machine Learning Algorithm. IJERT.

Džaferovic, E., & Hadžiabdic, K.K. (2021). Air Quality Prediction and Application Using Machine Learning Methods: A Case Study of Bjelave Neighborhood. Lecture Notes in Network Systems, 142.

Eslami, E., et al. (2019). A Real-Time Hourly Ozone Prediction System Using Deep Convolutional Neural Network. Neural Computing and Applications.

Feng, X., et al. (2015). Artificial Neural Network Forecasting of PM2.5 Pollution Using Air Mass Trajectory Based Geographic Model and Wavelet Transformation. Atmospheric Environment, 107, 118–128.

Goulier, L., et al. (2020). Modelling of Urban Air Pollutant Concentrations with Artificial Neural Networks Using Novel Input Variables. International Journal of Environmental Research and Public Health, 17(2025).

Ghazali, S., & Ismail, L.H. (2017). Air Quality Prediction Using Artificial Neural Networks. CORE.

Joharestani, M., et al. (2019). PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. Atmosphere, 10(373).

Kıymet Kaya & Öğüdücü, Ş.G. (2020). Deep Flexible Sequential Model for Air Pollution Forecasting. Scientific Reports, 10(3346).

Khanna, E. (2015). Neural Network Based Regression Testing. International Journal of Advanced Technology in Engineering and Science, 1(3).

Liang, Y.-C., et al. (2020). Machine Learning-Based Prediction of Air Quality. Applied Sciences, 10(9151).

Madhuri, V.M., et al. (2020). Air Pollution Prediction Using Machine Learning Supervised Learning Approach. IJSTR, 9(4), 118–123.

Mansour, S., Nasser, N., Karim, L., & Ali, A. (2014). Wireless Sensor Network-based Air Quality Monitoring System. IEEE Computing, Networking and Communications Conference.

Masih, A. (2019). Machine Learning Algorithms in Air Quality Modeling. Global Journal of Environmental Science and Management, 5(4), 515–534.

Memon, M. (2022). ANN vs CNN vs RNN: Neural Networks Guide. Levity.ai..

Mishra, A. (2018). Air Pollution Monitoring System Based on IoT: Forecasting and Predictive Modeling Using Machine Learning. AESPC 2018 Conference Proceedings.

Obodoeze, F. C., & Amadi, A. O. (2018). The Awka Smart City Project: Challenges and Roadmap. In Proceedings of the 2018 Biennial School of Engineering Conference, Unwana, Nigeria.

Pan, B. (2018). Application of XGBoost Algorithm in Hourly PM2.5 Concentration Prediction. IOP Conf. Ser.: Earth Environ. Sci., 113(012127).

Qadeer, K., et al. (2020). A Long Short-Term Memory Network for Hourly Estimation of PM2.5 Concentration in Two Cities of South Korea. Applied Sciences, 10(3984).

Rahimi, A. (2017). Short-Term Prediction of NO2 and NOx Concentrations Using Multilayer Perceptron Neural Network. Ecological Processes, 6(4).

Sankar, G.S., et al. (2019). Forecasting Air Quality Index Using an Ensemble of Artificial Neural Networks and Regression Models. Journal of Intelligent Systems, 28(5), 893–903.

Shah, S., et al. (2021). A Hybrid Model for Forecasting of Particulate Matter Concentrations Based on Multiscale Characterization and Machine Learning. Mathematical Biosciences and Engineering, 18(3), 1992–2009.

Shashi, R.G., et al. (2020). Analysis of Fine-Grained Air Quality and Prediction of Air Pollution. IRJET, 7(5), 6125–6127.

Shishegaran, A., et al. (2020). Prediction of Air Quality in Tehran by Developing the Nonlinear Ensemble Model. Journal of Cleaner Production, 259(120825).

Smartcitizenkit (2025). Smart Citizens Kit. Accessed online at https://docs.smartcitizen.me/hardware/kit/

Subramanian, V.N. (2016). Data Analysis for Predicting Air Pollutant Concentration in Smart City Uppsala. MSc Thesis, Uppsala Universitet.

Su, X., et al. (2020). Prediction of Ozone Hourly Concentrations by Support Vector Machine and Kernel Extreme Learning Machine. Atmospheric Pollution Research, 11(6), 51–60.

Wang, W., et al. (2019). Estimation of PM2.5 Concentrations in China Using a Spatial Back Propagation Neural Network. Scientific Reports, 9(13788).

Zamani, M., et al. (2019). PM2.5 Prediction Based on Random Forest, XGBoost and Deep Learning Using Multisource Remote Sensing Data. Atmosphere, 10(373).

Zhang, Y., et al. (2012). A review of air pollution prediction methods. Environment International, 32(6), 707–717.

Wikipedia Contributors. (2020). Awka. Retrieved from https://en.wikipedia.org/wiki/Awka

Published bimonthly by the Nigerian Institute of Electrical and Electronics Engineers (NIEEE) Awka Chapter, Nigeria..