FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 9, September, 2025, pp 218 – 226
DOI: https://doi.org/10.33003/fjs-2025-0909-3932

MACHINE LEARNING PREDICTION OF $PM_{2.5}$ IN LAGOS USING EMBEDDED REAL-TIME ENVIRONMENTAL MONITORING

[1,2]Odubanjo, Oduwole F., [2]Aluko, Tolulope O., [1]Sanni, Mohammed and [2]Jimoh, Oluwaseyi E.

[1]Department of Physics, University of Ilorin, Ilorin, Nigeria.
[2]Department of Physics Education, Federal College of Education (Technical), Akoka, Lagos, Nigeria.

*Corresponding authors' email: toluwisdom@gmail.com; tolulope.aluko@fcetakoka.edu.ng

ABSTRACT

This study aims to develop and evaluate an embedded system for real-time monitoring of PM2.5 and meteorological variables, with the goal of improving machine learning predictions of particulate matter concentrations in Lagos, Nigeria. Given the detrimental health effects of PM2.5, understanding its interaction with environmental factors is crucial for effective air quality control. Over two years (2021–2023), our innovative, custom-developed sensor system was deployed in Akoka, Lagos, to continuously and autonomously collected temperature, humidity, wind speed, atmospheric pressure, and PM2.5 data at two-minute intervals. Leveraging this robust data set, three machine learning algorithms: Random Forest (RF), Gradient Boosting (GB), and Support Vector Machine (SVM), were systematically evaluated for PM2.5 forecasting using R^2 and RMSE as performance metrics. The Random Forest model demonstrated the best performance ($R^2 = 0.77$; RMSE = 10.84 µg/m³), indicating high predictive capacity. Feature importance analysis revealed a limited impact of meteorological variables compared to unmeasured emission sources. This work demonstrates the feasibility of embedded real-time monitoring integrated with ML for urban air quality forecasting, supporting improved policy and public health strategies in rapidly urbanizing regions.

Keywords: Particulate Matter, Embedded System, Machine Learning, Air Quality Prediction, Meteorological Data

INTRODUCTION

The adverse impacts of particulate matter (PM) on human health and the environment make air quality a major global public health problem. PM is classified according to its aerodynamic diameter, with PM2.5 (particles with a diameter of 2.5 micrometers or less) and PM10 (particles with a diameter of 10 micrometers or less) being the most studied components. Particularly in urban areas like Lagos, Nigeria, PM2.5 poses significant health risks, as these fine particles can penetrate the bloodstream and lungs, leading to respiratory ailments, cardiovascular disorders, and premature mortality (Nathaniel & Xiaoli, 2020; Manisalidis et al., 2020; World Health Organization, 2021; Thangavel et al., 2022). Vehicle emissions, industrial discharges, and natural events like dust storms and wildfires are major sources of PM (McDuffie et al., 2021; Thangavel et al., 2022). Understanding the complex relationships between PM pollution and various meteorological factors including temperature, humidity, wind speed, and atmospheric pressure is essential for effective air quality management (Chen et al., 2020). Recognizing how these meteorological parameters interact with PM concentrations can enhance predictive models and inform regulatory actions. Hence, recent developments in sensor technology have enabled the creation of innovative embedded systems that gather environmental data in real-time. These devices can be deployed in various locations to continuously monitor air quality and weather conditions. By combining data from multiple sensors, researchers can build comprehensive databases that reflect the dynamic interplay between PM levels and meteorological conditions (Javaid et al., 2021). Machine learning (ML) has emerged as a powerful tool for analyzing large, complex datasets while traditional statistical methods often struggle to capture nonlinear relationships and interactions within the data. In contrast, ML algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB) have shown promise in improving prediction accuracy (Cha et al., 2021; Tran et al., 2024; Krishna et al., 2024). For example, SVM excels in classification and regression tasks, making it suitable for predicting PM concentrations based on meteorological parameters (Suárez-Sánchez et al., 2011). In cities like Lagos, where rapid population growth and industrialization exacerbate air pollution levels, accurate predictions of PM2.5 concentrations are crucial for enhancing air quality management. This can inform community awareness campaigns, regulatory guidelines, and public health advisories. This work introduces a novel embedded system designed to collect data on meteorological parameters and PM concentrations, with a specific focus on PM2.5. Despite advances in PM monitoring and ML modeling, few studies in sub-Saharan Africa, particularly in Lagos, have combined real-time embedded systems with comparative ML evaluations for PM2.5 prediction. This study aimed to evaluate the prediction capabilities of three machine learning algorithms: SVM, RF, and GB using this data. Specifically, the research addressed the following research questions: Which meteorological parameters most significantly influence PM2.5 in Lagos? Which ML model provides the most accurate predictions for real-time PM2.5 data collected via embedded systems? By filling this knowledge gap, this research aims to advance the understanding of air quality prediction models and their implications for environmental regulations and public health outcomes.

MATERIALS AND METHODS

Description of the Study Areas

The research was conducted in Akoka, a suburb of Lagos, Nigeria, which is notable for its high population density and prominent educational institutions, including the Federal College of Education (Technical). This region experiences a tropical climate characterized by distinct wet and dry seasons, which significantly impact $PM_{2.5}$ levels. The study area is located between Longitude N06º31'10" and N06º31'30" and Latitude E3º22′55″ and E3º23′5.3″, as illustrated in Fig. 1.

Exact sensor placement was at 3.5 meters above ground level and at GPS coordinates (6°31′20.0″ N, 3°23′00.0″ E) chosen to represent ambient neighborhood conditions consistent with WHO and national environmental agency guidelines.

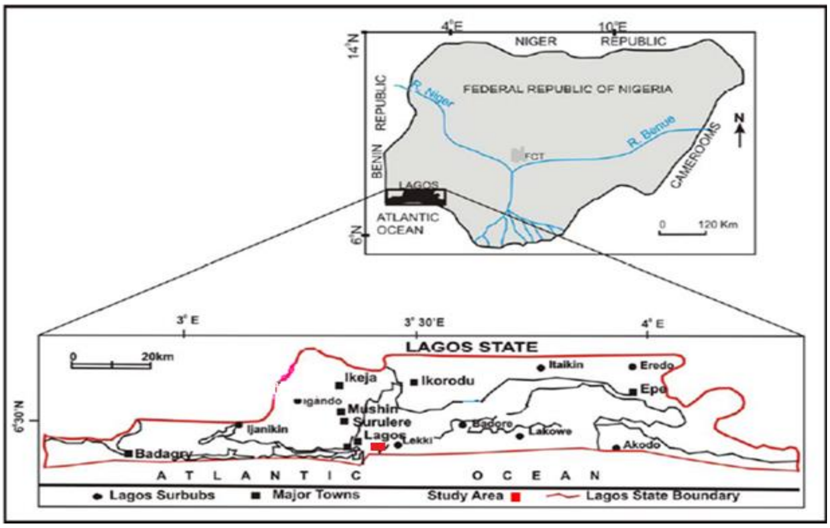

Figure 1: Map of Nigeria showing Lagos State and Federal College of Education (Technical) in Akoka

Instrument for Data Collection

The study employed a self-developed, easy-to-operate, and well-validated automated embedded system designed for real-time environmental monitoring of particulate matter (PM_1, $PM_{2.5}$, and PM_{10}) and key meteorological parameters. The system was purpose-built for autonomous, long-term outdoor operation in Lagos' urban environment, where reliability, resilience, and energy independence are essential. Sensor selection was informed by peer-reviewed validation studies and cost-effectiveness, prioritizing performance and suitability for extended urban deployments. The PMS7003 (Plantower) was used for particulate matter measurement, BMP180 for temperature and pressure, HIH4000 for humidity, and DFR-12 for wind speed. These sensors provide accurate, real-time environmental data and were chosen to minimize calibration drift and optimize system durability.

The system's power architecture was designed for off-grid, continuous operation, comprising a 60W solar panel to harvest solar energy, a charge controller to regulate voltage and protect the battery and circuitry, and a 40AH lithium-ion battery to store energy and support operation during low sunlight conditions or overnight. This robust power configuration ensures that all system components, sensors, controller, and communication units remain fully operational without external power, even in extended periods of poor weather or grid outages.

At the heart of the system are two integrated microcontrollers, each assigned distinct roles for modularity and performance. The ATSAMD21, a 32-bit ARM Cortex-M0+ microcontroller, handled real-time sensor sampling and data acquisition. The ESP8266-12E module, connected to the ATSAMD21 via UART, managed wireless data transmission to a remote station computer over Wi-Fi. The ESP8266 is a low-cost, low-power system-on-chip (SoC) with an integrated TCP/IP stack, making it ideal for Internet of Things (IoT) applications. By offloading network communication tasks to the ESP8266, the ATSAMD21 was able to focus fully on time-sensitive data acquisition and processing without network-induced latency, thereby improving overall system efficiency and stability.

Sensor interfacing was achieved using protocol-specific communication lines. The BMP180 (temperature and pressure) operates over I²C. The HIH4000 and DFR-12 (humidity and wind speed) provided analog outputs connected to the ADC channels of the ATSAMD21. The PMS7003 (particulate matter) communicated via UART, delivering digital output directly to the microcontroller. The PMS7003 sensor utilizes a laser scattering method based on Mie theory, wherein particles passing through a laser beam scatter light at various angles depending on size. The scattered light is analyzed to estimate concentrations of $PM_{1.0}$, $PM_{2.5}$, and PM_{10}, offering high-precision particulate measurement in real time.

A software was developed using Embedded C in the Arduino IDE for the ATSAMD21, which managed sensor data acquisition and time stamping. A Python script executed on the ESP8266-12E coordinated wireless data transfer to the station computer via Wi-Fi, where all received data were logged and stored for subsequent analysis. This setup provided real-time monitoring and centralized storage while relying on the station computer for persistent data logging. The system block diagram, design components' flowchart, and the exterior and interior views of the installed device are shown in Figure 2, 3 and 4 confirming both the technical implementation and real-world usability of the designed embedded system.

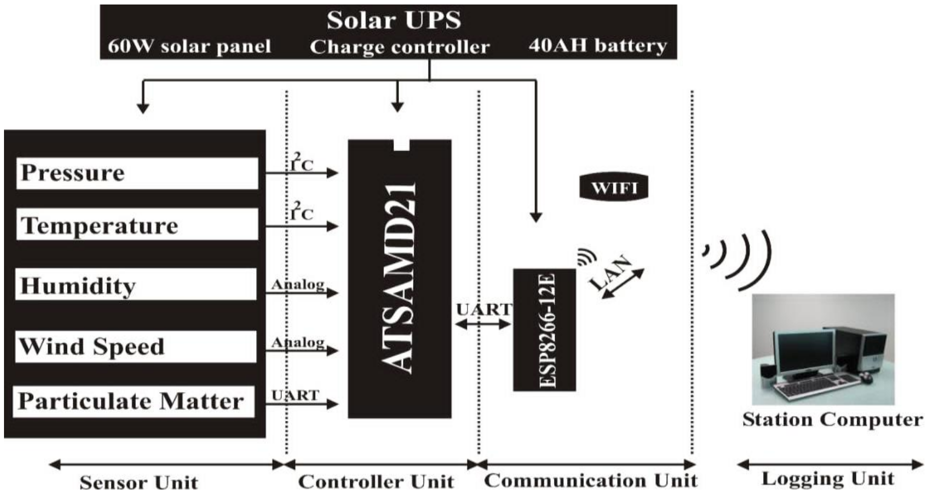

Figure 2: Block Diagram of the Embedded System

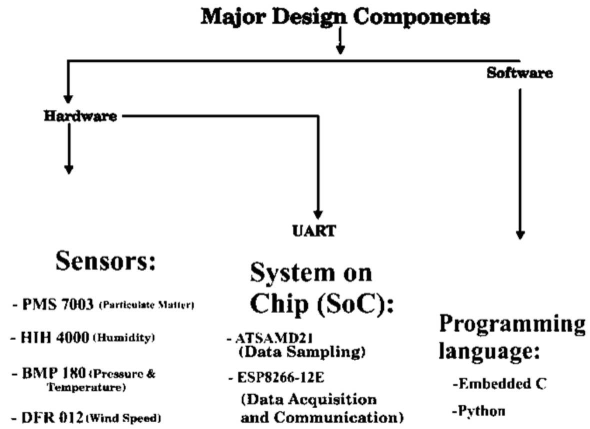

Figure 3: Flowchart of the Design Components

Figure 4: Exterior and Interior View of the Installed Device

Calibration and Validation of the Designed Instrument

Calibration and validation were undertaken to ensure the reliability and accuracy of the system's measurements in comparison to reference-grade instruments. The system was calibrated through a co-location exercise with certified meteorological instruments from the Department of Geography, University of Lagos, and a DM-106A handheld particulate matter monitor over a continuous 30-day period. During this period, data from the developed embedded system were simultaneously collected alongside standard instruments, capturing a broad range of ambient conditions, including morning and evening rush hours, midday heat, and varying humidity and wind profiles.

Discrepancies between the system readings and those of the reference devices were statistically analyzed to determine appropriate correction factors for each parameter. These correction algorithms were implemented within the firmware of the ATSAMD21 microcontroller to automatically adjust all real-time measurements at the point of acquisition. This

firmware-level calibration ensured that future data collected would reflect corrected values without post-processing.
Sensor mounting and orientation were optimized according to international best practices: placement at 3.5 meters above ground, away from direct emission sources, and shaded to prevent solar heating artifacts. Routine maintenance was conducted weekly to clean sensor inlets and ensure consistent performance throughout the calibration and validation period.
Following the implementation of calibration adjustments, a three-month validation phase was carried out from April to June 2019. During this period, the developed instrument remained co-located with reference-grade instruments, and its data were compared to standard measurements through time series plots and regression analysis. Validation results were quantitatively assessed using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2). The summarized results are presented in Table 3.1 and indicate a strong correlation and low error margins across all parameters.

Table 1: Summary of The Instrument Validation Results

Parameters	Slope	Intercept	R	R^2	RMSE	MAE
Pressure	0.91	95.81	0.93	0.86	2.94	2.67
Temperature	1.01	1.10	0.96	0.92	1.52	1.38
RH	1.06	-5.75	0.96	0.92	1.95	1.37
Wind Speed	0.89	0.92	0.96	0.92	0.52	0.23
PM_1	0.85	6.83	0.93	0.86	3.55	2.88
$PM_{2.5}$	1.04	5.4	0.96	0.92	4.33	2.48
PM_{10}	1.07	6.29	0.94	0.88	4.91	1.92

These results confirm that the developed instrument meets international benchmarks for continuous air quality monitoring.

Method of Data Collection

Data were collected over a two-year period from May 1, 2021, to April 30, 2023, at a site located in the Federal College of Education (Technical), Akoka, Lagos. The embedded system captured PM concentrations (PM_1, $PM_{2.5}$, PM_{10}) and meteorological parameters (temperature, humidity, pressure, wind speed) every two minutes. These readings were averaged into hourly, daily, and monthly intervals to support various analyses and model training. Temporal coverage and high-frequency sampling (two-minute interval) ensured weather and emission events were accurately captured across diurnal and seasonal cycles.

Data Preprocessing

All data preprocessing and modeling were implemented using Python 3.8 in a Jupyter Notebook environment, with key libraries including Pandas 1.1.5, NumPy 1.19.5, and Scikit-learn 0.24.1. Outliers in the dataset were addressed using the Interquartile Range (IQR) method; data points falling outside the range defined by Q1 – 1.5×IQR and Q3 + 1.5×IQR were excluded to prevent their influence on model training and to ensure robust performance.
To handle missing data, Little's MCAR test was conducted to confirm that missingness occurred at random. Test results indicated the data were missing completely at random (MCAR), justifying the use of mean imputation for its simplicity and minimal distortion to underlying patterns. For model development, the dataset was split into 80% for training and 20% for testing, ensuring temporal continuity was preserved throughout. All evaluation metrics, including those for validation and performance comparison, were computed from the held-out test set to maintain an unbiased assessment of model generalization.

Machine Learning Algorithms

This study employed three machine learning algorithms: Random Forest (RF)**,** Gradient Boosting (GB)**,** and Support Vector Machine (SVM)**,** to predict $PM_{2.5}$ concentrations in Lagos. These algorithms were selected based on their proven effectiveness in capturing complex patterns within environmental datasets and their capacity to handle non-linear relationships, noisy data, and multivariate inputs.
Random Forest was chosen for its ensemble learning approach, which constructs multiple decision trees during training and aggregates their outputs. This mechanism effectively reduces overfitting and increases model stability, making it especially suitable for modeling the complex interactions among meteorological variables that influence air quality.
Gradient Boosting**,** another ensemble method, builds decision trees sequentially, where each new tree corrects the errors made by the previous one. Its additive learning framework allows it to model intricate non-linear relationships. It is widely known for delivering high predictive accuracy in environmental and atmospheric applications.
Support Vector Machine was incorporated due to its capacity to handle high-dimensional datasets and its ability to find optimal hyperplanes for classification and regression tasks. With the use of appropriate kernel functions, SVM is capable of handling both linear and non-linear relationships, which is advantageous in air pollution modeling where interactions among features are often complex and non-linear.

Model Selection, Training, and Optimization

Following algorithm selection, a rigorous workflow was implemented to train, optimize, and evaluate the models. The process began with data preprocessing and was followed by model training using an 80/20 split for training and testing. To enhance reliability and avoid overfitting, K-fold cross-validation was applied, where the dataset was partitioned into five folds and iteratively trained and validated across all subsets. This approach ensures robust performance estimation and generalizability, flowchart shown in Figure 5.
To optimize each model's performance, GridSearchCV from the Scikit-learn library was employed for hyperparameter tuning. This method performs an exhaustive search over specified parameter combinations to identify the best configuration. The search was conducted using 5-fold cross-validation integrated within GridSearchCV, ensuring that hyperparameters were not only optimized for the training set but also validated consistently across multiple data partitions. For the Random Forest model, the final configuration used 100 decision trees and a maximum depth of 10. The Gradient Boosting model employed 200 estimators and a learning rate of 0.1. The SVM model was optimized for kernel type and

regularization parameter. These hyperparameter choices were carefully selected to balance model complexity, computation time, and predictive performance.

Each trained model was evaluated on the testing dataset using three core performance metrics: Root Mean Square Error (RMSE)**,** Mean Absolute Error (MAE)**,** and R-squared (R^2)**.** These metrics allowed for a quantitative comparison of the models' predictive abilities and informed the final model **selection**, which favored Random Forest based on its superior balance of accuracy and generalization, as elaborated in the Results and Discussion section.

Despite the robust methodology, some limitations were acknowledged. The study's temporal scope, limited to two years, may not fully capture long-term or seasonal variations in $PM_{2.5}$ concentrations, potentially affecting model generalizability. Moreover, mean imputation was used to handle missing values under the assumption of randomness, which, if violated, could introduce bias. Lastly, while cross-validation and hyperparameter tuning strengthened model reliability, they also introduced substantial computational demands.

Nevertheless, this modeling pipeline demonstrates a rigorous and scalable framework for real-time air quality prediction. It underscores the value of combining embedded sensing systems with advanced machine learning techniques to inform public health policies and urban air quality management.

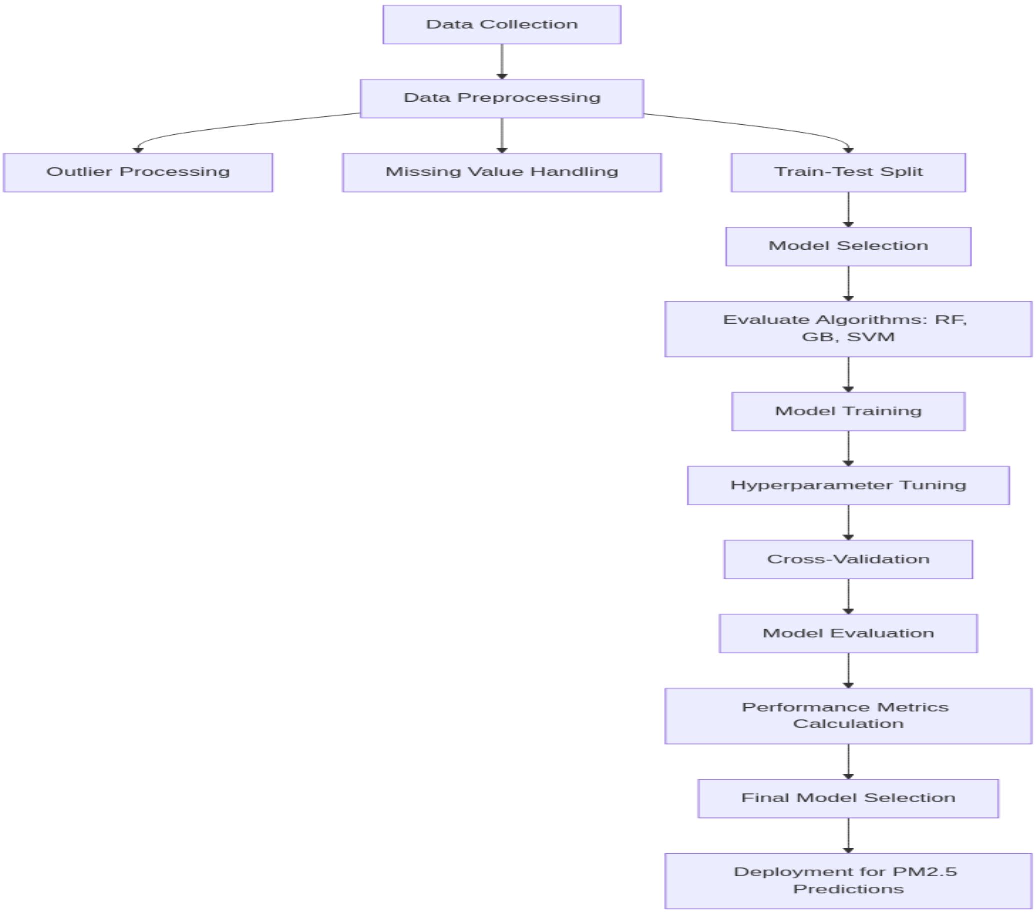

Figure 5: Machine Learning Flow Chart

RESULTS AND DISCUSSION

A. $PM_{2.5}$ Average Daily and Average Seasonal Concentration Levels

Table 2: $PM_{2.5}$ Average Daily and Average Seasonal Concentration Levels

Statistic	Value (µg/m³)
Minimum Value	0.94
Maximum Value	86.42
Range	85.94
Mean Value	37.39
Standard Deviation	26.60
Seasonal Mean (Dry)	37.48
Seasonal Mean (Rainy)	31.83

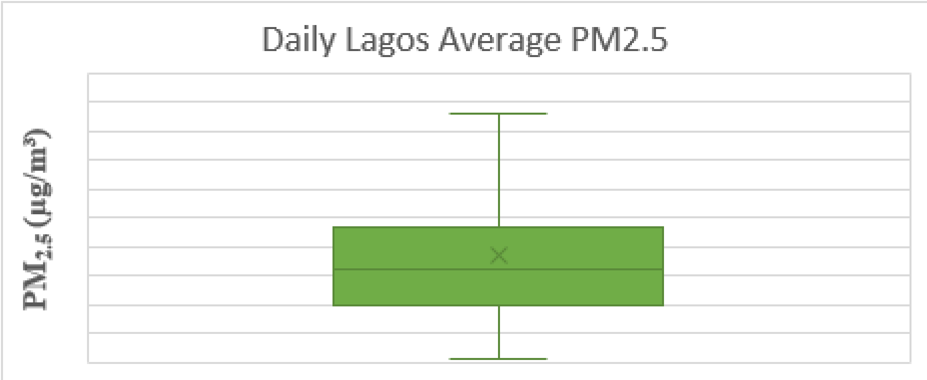

Figure 6: Boxplots Representing Daily Average of $PM_{2.5}$ Concentrations

Lagos, Nigeria, is characterized by persistently high $PM_{2.5}$ concentrations, as demonstrated by the analysis in Table 2 and Figure 6. The daily $PM_{2.5}$ levels fluctuate substantially, with values ranging from 0.94 µg/m³ to 86.42 µg/m³, and a mean daily concentration of 37.39 µg/m³, well above recommended air quality guidelines (World Health Organization, 2021). This finding is consistent with recent work by Yahaya et al. (2023), further substantiating the ongoing air quality challenges faced by Lagos.

Seasonally, dry periods exhibit a slightly higher mean $PM_{2.5}$ (37.48 µg/m³) than rainy periods (31.83 µg/m³), which reinforces global trends of improved air quality during rainfall months due to the natural cleansing effect of precipitation (Shukla *et al.,* 2008; Lala *et al.,* 2023). The high standard deviation (26.60 µg/m³) highlights significant variability, likely driven by episodic events such as traffic build-ups and intermittent industrial emissions. Such patterns are illustrated by the interquartile range and outliers depicted in Figure 4's boxplot.

These observations align with the pattern of elevated $PM_{2.5}$ prevalent in rapidly urbanizing cities of the developing world (Okimiji *et al.,* 2021; Obike–Martins *et al.,* 2022; Atou *et al.,* 2022; Okudo *et al.,* 2022; Emekwuru *et al.,* 2023; Odubanjo *et al.,* 2024). Major sources in such environments include rapid urban expansion, dense traffic, industrial activity, and frequent combustion of fossil fuels. Given the observed levels and variances, the public health implications are severe, particularly concerning respiratory and cardiovascular health burdens (World Health Organization, 2021; Manisalidis *et al.,* 2020).

These results directly address the first research question, highlighting not just the magnitude but also the variability of $PM_{2.5}$ exposures in Lagos. Mitigating these risks will necessitate continuous monitoring, detailed source apportionment, and targeted interventions.

Machine Learning Model Prediction for $PM_{2.5}$ in Lagos

Three machine learning models: Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB), were robustly evaluated using key prediction metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). The results are summarized in Table 3.

Table 3: Prediction Performance of RF, SVM, and GB Models for $PM_{2.5}$ in Lagos

Model	RMSE	MAE	R-Squared
Random Forest	10.84	6.62	0.77
Support Vector Machine	22.81	17.79	0.24
Gradient Boosting	11.89	7.73	0.72

The Random Forest model achieved the best predictive performance, explaining 77% of the variance in $PM_{2.5}$ levels (R^2 = 0.77, RMSE = 10.84). This outperformance is attributable to RF's capacity to model complex, nonlinear relationships through ensemble averaging (Ameer *et al.,* 2019; Fang *et al.,* 2021). In contrast, SVM's poor results (R^2 = 0.24, highest RMSE and MAE) suggest its limitations—possibly related to kernel or scaling sensitivity and data complexities typical in urban air environments (Rodríguez-Pérez *et al.,* 2022). The GB model provided intermediate results and, like RF, leveraged tree-based learning to achieve substantial predictive accuracy (R^2 = 0.72).

These findings are also consistent with prior air quality modeling studies in similar contexts (Gupta *et al.,* 2023; Mahmud et al., 2022; Vignesh *et al.,* 2023). Thus, in direct response to the second research question, RF is confirmed as the preferred algorithm for this embedded system application.

PM2.5 Prediction Visualization and Residual Diagnostics

Figure 7 shows a combined scatter and line plot of actual and RF-predicted $PM_{2.5}$ values. Blue dots denote observed, orange dots predicted, and smoothed trend lines for both are also presented.

Visual inspection of Figure 5 reveals that RF effectively captures temporal variation in $PM_{2.5}$, though brief divergences during extreme pollution events are apparent. Residual diagnostics (Figures 8 and 9) indicate that prediction errors (actual minus predicted) cluster around zero but show heavier tails and greater variance during episodes of high $PM_{2.5}$.

A Shapiro-Wilk test ($p < 0.05$) confirmed the non-normality of residuals, indicating persistent challenges in accurately forecasting rare or extreme pollution spikes. This is a known limitation in environmental time series where underlying drivers (e.g., sudden traffic surges, industrial releases) are not directly measured.

Further, the residual distribution pattern underscores a need for hybrid/ensemble models or error forecasting frameworks for improved handling of such variability, as recommended in the recent literature.

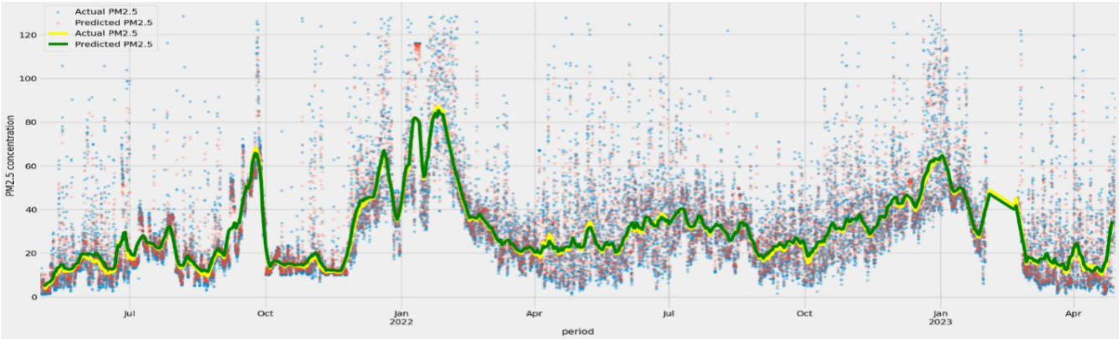

Figure 7: Visualization of scatter plot and line Graph of actual $PM_{2.5}$ and Predicted $PM_{2.5}$ for the RF Algorithm in Lagos

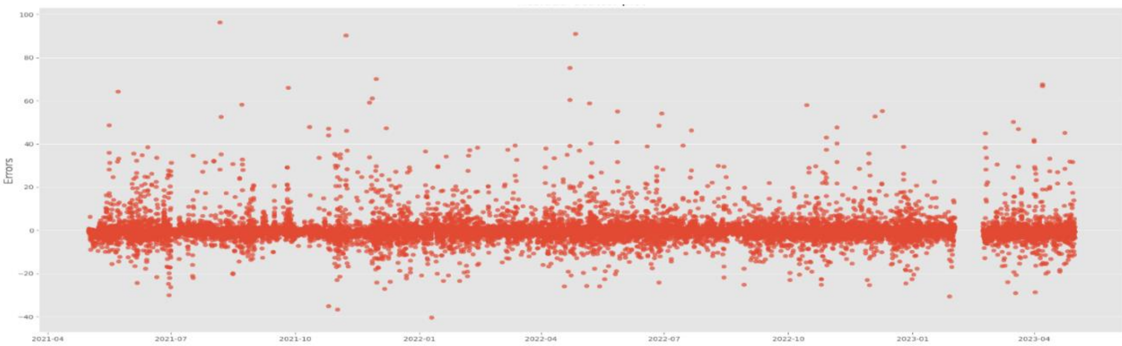

Figure 8: $PM_{2.5}$ Residual Scatter Plots

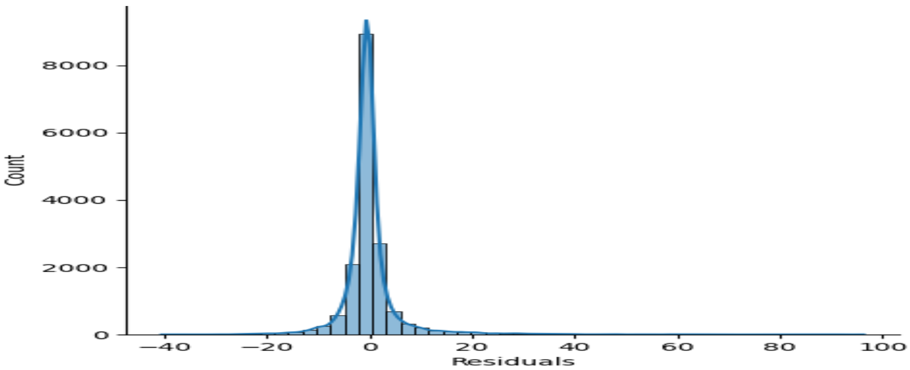

Figure 9: $PM_{2.5}$ Residual Distribution Plot

Feature Importance Analysis for RF Model

Table 4: Feature Importance Analysis for the Random Forest Algorithm in $PM_{2.5}$ Prediction

Input Variables	Importance (%)
Humidity	10.92
Temperature	9.38
Wind Speed	8.24
Pressure	5.86

Meteorological conditions contribute only partially to explaining $PM_{2.5}$ variability. This finding aligns with previous research in tropical megacities, which highlights that air pollution levels are shaped more by local, often unmeasured sources such as traffic, industrial activity, and open burning than by meteorological dynamics alone (Theogene *et al.,* 2020; Khalis *et al.,* 2022). The analysis further indicates that relying solely on meteorological measurements may overlook rapid changes in local emissions, such as those caused by spikes in traffic or short-term combustion events, which standard weather data are unlikely to capture. Incorporating additional data types, such as real-time information on traffic volumes, land use, or emission inventories, could provide a more comprehensive basis for predicting $PM_{2.5}$ concentrations.

The feature importance assessment directly addresses the first research question by showing that, while meteorological variables do have an effect on $PM_{2.5}$ levels in Lagos, their impact is limited compared to that of unmeasured local emission sources. This underscores the need for future modeling efforts to integrate direct measurements of anthropogenic activities to better capture the underlying drivers of air pollution in the city.

CONCLUSION

This study developed and systematically evaluated a real-time embedded system for monitoring PM2.5 and meteorological variables in Lagos, Nigeria, a city confronting severe and persistent air quality challenges. By combining cost-effective sensor technologies with ML-based forecasting, the work presents a replicable framework for high-frequency, continuous pollution monitoring in resource-constrained settings.

Among the tested ML models, Random Forest emerged as the most accurate (RMSE = 10.84, R^2 = 0.77), validating the strength of ensemble learning for complex, noisy environmental datasets. Feature analysis confirmed that meteorological factors, while significant, do not solely determine $PM_{2.5}$ dynamics in urban Lagos; substantial variance remains unaccounted for, likely due to non-meteorological emissions not measured here. This underscores the urgent need for future studies to integrate traffic, industrial, and other anthropogenic activity data for greater explanatory and forecasting power.

The present analysis did not employ formal statistical hypothesis testing (e.g., ANOVA, Kruskal-Wallis) to evaluate model differences, and did not generate confidence intervals for error metrics. Addressing these in future work will enhance model interpretability and reliability. Further, generalizability to other settings—beyond Lagos or across various West African cities—remains an important research direction.

Despite these acknowledged limitations, this work provides an important foundation for scalable, intelligent air quality management. By enabling granular, real-time, and predictive air quality assessment, the proposed system can inform public health planning, early warning, and regulatory action in rapidly developing urban centers.

REFERENCES

Ameer, S., Shah, M., Khan, A., Song, H., Maple, C., Islam, S., & Asghar, M. (2019). Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE Access*, *7*, 1-1. https://doi.org/10.1109/ACCESS.2019.2925082

Atoui, A., Slim, K., Andaloussi, S.A., Moilleron, R., & Khraibani, Z. (2022). Time Series Analysis and Forecasting of the Air Quality Index of Atmospheric Air Pollutants in Zahleh, Lebanon. *Atmospheric and Climate Sciences*, 12, 728- 749. https://doi.org/10.4236/acs.2022.124040

Cha, G. W., Moon, H. J., & Kim, Y. C. (2021). Comparison of random forest and gradient boosting machine models for predicting demolition waste based on small datasets and categorical variables. *International Journal of Environmental Research and Public Health, 18*(16), 8530. https://doi.org/10.3390/ijerph18168530

Chen, Z., Chen, D., Zhao, C., Kwan, M.-P., Cai, J., Zhuang, Y., Zhao, B., Wang, X., Chen, B., Yang, J., Li, R., He, B., Gao, B., Wang, K., & Xu, B. (2020). On PM2.5 concentrations across China: A review of methodology and mechanism. *Environmental International, 145,* 105558. https://doi.org/10.1016/j.envint.2020.105558

Emekwuru, N., & Ejohwomu, O. (2023). Temperature, humidity and air pollution relationships during a period of rainy and dry seasons in Lagos, West Africa. *Climate, 11*(5), 113. https://doi.org/10.3390/cli11050113

Fang, Z., Yang, H., Li, C., Cheng, L., Zhao, M., & Xie, C. (2021). Prediction of PM2.5 hourly concentrations in Beijing based on a machine learning algorithm and ground-based LiDAR. Archives of Environmental Protection, 47(3), 25-37. https://doi.org/10.24425/aep.2021.138468

Gupta, N.S., Yashvi, M., Khyati, H., Raahil, A., and Valarmathi, B. (2023). Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis. *Journal of Environmental and Public Health*, (2023)1. https://doi.org/10.1155/2023/4916267

Javaid, M., Haleem, A., Rab, S., Singh, R. P., & Suman, R. (2021). Sensors for daily life: A review. Sensors International, 2, 100121. https://doi.org/10.1016/j.sintl.2021.100121

Khalis, M., Toure, A.B., El Badisy, I., Khomsi, K., Najmi, H., Bouaddi, O., Marfak, A., Al-Delaimy, W.K., Berraho, M., & Nejjari, C. (2022). Relationship between Meteorological and Air Quality Parameters and COVID-19 in Casablanca Region, Morocco. *International Journal of Experimental Research and Public Health*, 19(9),4989.https://doi.org/10.3390/ijerph19094989

Krishna, R., Jayanthi, D., Sam, D. S., Kavitha, K., Maurya, N. K., & Benil, T. (2024). Application of machine learning techniques for churn prediction in the telecom business. *Results in Engineering, 24,* 103165. https://doi.org/10.1016/j.rineng.2024.103165

Lala, M.A., Onwunzo, C.S., Adesina,O.A & Sonibare, J.A.(2023). Particulate matter pollution in selected areas of Nigeria: Spatial analysis and risk assessment. *Case Studies in Chemical and Environmental Engineering, 7 (2023).* https://doi.org/10.1016/j.cscee.2022.100288.

Mahmud, S., Ridi, T.B.I., Miah, M.S., Sarowar, S., & Elahee, S. (2022). Implementing Machine Learning Algorithms to Predict Particulate Matter (PM2.5): A Case Study in the Paso del Norte Region. *Atmosphere,*13(12),2100. https://doi.org/10.3390/atmos13122100

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontier in Public Health,* 8 (14). https://doi.org/10.3389/fpubh.2020.00014

McDuffie, E. E., Martin, R. V., Spadaro, J. V., Burnett, R., Smith, S. J., Hammer, M. S., Van Donkelaar, A., Bindle, L., Shah, V., Jaeglé, L., Luo, G., Yu, F., Adeniran, J. A., Lin, J., & Brauer, M. (2021). Source sector and fuel contributions to ambient PM2.5 and attributable mortality across multiple spatial scales. *Nature Communications*, 12(1), 1-12. https://doi.org/10.1038/s41467-021-23853-y

Nathaniel, M. W., & Xiaoli. (2020). Air Quality Levels and Health Risk Assessment of Particulate Matters in Abuja Municipal Area, Nigeria. *Atmosphere.* 11(18), 817. https://doi.org/10.3390/atmos11080817

Obike-Martins, V., Fashola, M. S., Oji, N. C., Nwaugo, V. O., Uchenna, F. U., Okwy-Irokanulo,I N., & Enya, E. (2022). Effects of Seasonal Variation on Air Quality and Microbial Composition around Sawmilling Sites in South East Nigeria. *Journal of Advances in Microbiology*, *22*(6), 54–62.https://doi.org/10.9734/jamb/2022/v22i630469

Odubanjo, O. F., Falaiye, O. A., Orosun, M. M., & Sanni, M. (2024). Investigation of particulate matter Air Quality Index (AQI) and risk assessment in some locations in Nigeria. *Journal of the Nigerian Society of Physical Sciences*, *6*(4). https://doi.org/10.46481/jnsps.2024.2120

Okimiji, O.P., Techato, K., Simon, J.N., Tope-Ajayi, O.O., Okafor, A.T., Aborisade, M.A., & Phoungthong, K. (2021). Spatial Pattern of Air Pollutant Concentrations and their Relationship with Meteorological Parameters in Coastal Slum Settlements of Lagos, Southwestern Nigeria. *Atmosphere*, *12(11)*, 1426. http://dx.doi.org/10.3390/atmos12111426

Okudo, C. C., Ekere, N. R., & Okoye, C. O. B. (2022). Evaluation of particulate matter ($PM_{2.5}$ and PM_{10}) concentrations in the dry and wet seasons as indices of air quality in Enugu urban, Enugu State, Nigeria. Journal of Chemical Society of Nigeria, 47(5). https://doi.org/10.46602/jcsn.v47i5.807

Rodríguez-Pérez, R., & Bajorath, J. (2022). Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. Journal of Computer-Aided Molecular Design, 36(5), 355-362. https://doi.org/10.1007/s10822-022-00442-9

Shukla, J., Misra, A. K., Sundar, S., & Naresh, R. (2008). Effect of rain on removal of a gaseous pollutant and two different particulate matters from the atmosphere of a city. Mathematical and Computer Modelling, 48(5-6), 832-844. https://doi.org/10.1016/j.mcm.2007.10.016

Suárez Sánchez, A., García Nieto, P., Riesgo Fernández, P., Del Coz Díaz, J., & Iglesias-Rodríguez, F. (2011). Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). *Mathematical and Computer Modelling, 54*(5-6), 1453-1466. https://doi.org/10.1016/j.mcm.2011.04.017

Thangavel, P., Park, D., & Lee, Y. C. (2022). Recent insights into particulate matter (PM2.5)-mediated toxicity in humans: An overview. *International Journal of Environmental Research and Public Health, 19*(12), 7511. https://doi.org/10.3390/ijerph19127511

Theogene, I., Abubakar, S., & Jean, P. N. (2020). Establishing a Relationship between Meteorological Parameters and Criteria Pollutant Concentration in Delhi. *International Journal of Science and Research Methodology.* Vol. 15 (1), 30-44. https://ssrn.com/abstract=3800806

Tran, V. A., Khuc, T. D., Truong, X. Q., Nguyen, A. B., & Phi, T. T. (2024). Application of potential machine learning models in landslide susceptibility assessment: A case study of Van Yen district, Yen Bai province, Vietnam. *Quaternary Science Advances, 14,* 100181. https://doi.org/10.1016/j.qsa.2024.100181

Vignesh, P. P., Jiang, J. H., & Kishore, P. (2023). Predicting $PM_{2.5}$ concentrations across USA using machine learning. *Earth and Space Science,* 10 (10). https://doi.org/10.1029/2023EA002911

World Health Organization. (2021). WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization. https://apps.who.int/iris/handle/10665/345329

Yahaya, T., & Abdulazeez, A. (2023). Concentrations and health risks of particulate matter ($PM_{2.5}$) and associated elements in the ambient air of Lagos, Southwestern Nigeria. *Borno Researcher*, *21*(3), 2141-2149. https://doi.org/10.4314/br.v21i3.9