IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

IoT-Based Smart Framework for Air Quality Prediction

Ishwari Pawar, Sanika Mahajan, Tirtha Salunkhe, Dr. Rahul Agrawal
Electrical Engineering
Guru Gobind Singh College of Engineering and Research Centre

Abstract: *Air pollution levels in busy traffic areas vary depending on the number of vehicles, road conditions, and weather.Fixed pollution monitoring stations may not give detailed, local information.This paper introduces an Internet of Things (IoT)-based framework that uses distributed pollution sensors, environmental data, cloud services, and machine learning to predict air quality in the near future.The framework is built from four studies on IoT sensing, portable urban monitoring, edge-based PM2.5 forecasting, and an SV-CNN ensemble combined with K-nearest-neighbour (KNN) imputation.The sensor nodes gather data on particulate matter, and in some cases, gases, temperature, humidity, and weather.A microcontroller sends time-stamped measurements to a cloud platform, where data is checked for accuracy, missing values are handled, features are prepared, and predictions are made.A regression model can estimate pollutant concentration or the Air Quality Index (AQI). A classifier can determine the air quality category, but an edge model can offer quicker local predictions when there is limited internet connection.The results are shown on a dashboard that includes location-based alerts.Research findings from the original studies show the effectiveness of random-forest regression, hybrid NARX models, CNN-Bi-LSTM forecasting, and SV-CNN classification.However, these results are based on different datasets and tasks and are not considered as part of the experiments for the framework introduced in this paper.The paper presents a realistic design and testing plan for use near traffic lights, schools, hospitals, and other areas where people are exposed to high levels of pollution.*

Keywords: Internet of Things, air quality prediction, traffic congestion, air quality index, machine learning, edge computing, sensor networks

I. INTRODUCTION

Air pollution remains a ongoing problem in cities because factors like car emissions, dust from roads, industrial operations, and weather changes can all mix together to create pollution levels that vary a lot.Common pollutants that are tracked in monitoring systems are PM2.5, PM10, carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3).Fine particulate matter is particularly concerning because it can go deep into the lungs.At busy traffic intersections, pollution levels can spike, but these spikes may not be captured by a few fixed monitoring stations.While traditional monitoring stations can give accurate data, the cost of setting them up and keeping them running can limit how widely they are spread out.Using IoT-based sensors is a useful alternative: Relatively small sensor nodes can gather data from multiple locations and send it to a central service for analysis.Kumar and Doss introduced AIRO, a portable IoT monitoring system designed for use in urban areas, which uses cloud-based data and connects through a smartphone interface [1].Moursi and their team combined IoT, machine learning, and edge computing to predict PM2.5 levels and tested the approach on both a standard computer and a Raspberry Pi [2].Banciu et al.showcased a cloud-connected sensor monitoring system and compared different regression models for predicting air quality index (AQI) [3].Alnowaiser et al.suggested using an SV-CNN ensemble along with KNN imputation for air-quality classification [4].Building on these various approaches, this paper introduces a unified framework specifically

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

for areas with heavy traffic.The aim is not to present a new model that has been tested, but to establish a design that integrates sensing, data quality management, prediction, and alerts for users.

The framework can handle both numerical predictions and estimates of air-quality categories, but this depends on the availability and quality of the sensor data..

II. RELATED WORK

A. Portable IoT Monitoring for Urban Areas

Kumar and Doss created AIRO, a decentralized system for monitoring air quality.It is designed to give real-time data on air quality index (AQI) at any selected place.Their initial version had features like GPS and Wi-Fi, along with sensors for measuring particles and gases, a screen for displaying results, and a way to connect to the cloud.The research also suggested using a CNN-Bi-LSTM model to predict daily and weekly AQI levels for an entire city, and it included a mobile app for checking measurements and forecasts [1].The portable nature of AIRO makes it useful in areas with heavy traffic, as it can gather data from places where fixed monitoring stations are not available.However, one challenge is that the device and the prediction model together form a particular setup and process; therefore, their performance needs to be tested and confirmed in each new environment where they are used.

B. Edge-Based PM2.5 Prediction

Moursi and their team suggested an IoT-based system that uses a nonlinear autoregressive model with external inputs (NARX) along with several machine learning methods to predict PM2.5 levels for the next hour.They took into account recent pollution data and weather information.They tested running the system on a regular computer and also on a Raspberry Pi 4.Their findings show that using NARX with LSTM offered very good prediction results, while combining NARX with XGBRF gave a good mix of accuracy and speed on the Raspberry Pi [2].This suggests a mixed approach that uses the cloud for more detailed analysis and local processing for when internet connection is not reliable.

C. IoT Sensors and Regression for AQI

Banciu and their team developed an IoT device that monitors environmental conditions, sends the collected data to ThingSpeak, and uses a TensorFlow-based regression model to predict air quality index levels.Their dataset included 5,869 data points across six different parameters.They tested various neural network setups against random forest regression and found that a random forest model with 100 estimators gave the best results in terms of mean absolute error.The model achieved an MAE of 0.2785 for AQI 10 and 0.2483 for AQI 2.5.These results are specific to the dataset and definitions used in their study and may not directly apply to real-world traffic site applications.

D. Ensemble Classification and Missing-Value Handling

Alnowaiser and their team developed an SV-CNN stacked ensemble model that combines a support vector machine with a convolutional neural network, using soft voting to make the final classification decision.They applied KNN imputation to address missing data and achieved high performance metrics in their classification experiment, including 99.17% accuracy, 97.75% precision, 95.24% recall, and a 96.52% F1 score [4].In another test using a Beijing PM2.5 dataset, they reached 99.38% accuracy using the same KNN imputation method.These results are based on the specific data and setup used in their study and cannot be directly compared to regression error measures or hourly forecasting performance.The research highlights four key design ideas: place sensors in relevant locations, incorporate weather and time-related information, handle missing data properly, and choose the right model based on the task and available hardware.A practical framework should present its results on a clearly explained dataset before making any claims about its accuracy.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

III. PROPOSED SYSTEM ARCHITECTURE

The framework is divided into seven steps: first, collecting sensor data, then checking it with the microcontroller and locally, followed by sending the data through the internet or a gateway.Next, the data is stored in the cloud, then processed, used for machine learning predictions, and finally displayed on a dashboard with alerts.This setup can be used at traffic lights, intersections, school zones, hospital areas, and other places where it's important to monitor activity.The flow goes like this: Sensors → Microcontroller → Network/Gateway → Cloud Database → Data Processing → ML Prediction → Dashboard/Alerts.

A. Sensing Layer

Each node collects time-stamped data from the sensors included in its hardware setup.The main setup might have PM2.5 and PM10 sensors, temperature and humidity sensors, and optional gas sensors for carbon monoxide, nitrogen dioxide, sulfur dioxide, or ozone.If there are dependable sources, traffic levels and weather details can also be included.A GPS system or a set location code links each data point to a specific place.Keeping the sensors properly calibrated, designing the housing well, and placing them away from direct sunlight or obstacles help ensure accurate and consistent results.

B. Communication and Storage Layer

A microcontroller or single-board computer does simple checks to ensure data is within expected ranges, adds time stamps, and sends the data via Wi-Fi, a cellular network, or a nearby gateway.The cloud service keeps the original data and the processed data in separate places, which helps track where each piece of information came from.If the connection is lost, the device can save the data locally and send it later once the connection is back.The system continues to work smoothly in such situations

C. Data Preparation

Incoming data is checked for missing fields, duplicates, unreasonable values, inconsistent units, and time gaps.Missing data shouldn't be filled automatically.Short gaps can be handled with a proper imputation method, but long gaps or sensor issues should be noted.KNN imputation is one method suggested in [4], but it should be compared with simpler options and used only on training data when building the model to prevent data leakage.If the model requires it, numerical data can be normalized.For forecasting, it's better to split the data into training, validation, and test sets based on time, as random splitting might include future data in the training set.

D. Prediction Layer

The framework divides the prediction process into two separate tasks.Regression is used to estimate future levels of pollutants or numeric targets related to air quality indexes (AQI).Some models that can be used for this include random forest regression, neural networks, and temporal models like NARX or LSTM [2], [3].Classification, on the other hand, is used to assign a category, such as an AQI range.When there is labeled data and the necessary tools are available, models like SV-CNN can be tested [4].The best choice of model depends on factors like the time frame of the prediction, the amount of data available, how easy the model is to understand, the computing power needed, and the cost of missing high-pollution events.The framework does not assume that one model will work best for every location.

E. Dashboard and Alert Layer

The dashboard shows current measurements, predicted values, the time the data was collected, where the sensor is located, and the quality of the data.Alerts are sent out if a validated limit is crossed or if the model predicts a high-pollution level.Each alert should include the location, the exact time, the pollutant or index involved, and whether the value is based on actual measurements or a prediction.The interface also displays recent trends and a map of all the sensors, helping users and local authorities spot areas with possible pollution issues.Any health-related advice should follow local guidelines, not just what the model suggests.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

IV. METHODOLOGY

A. Data Collection Plan

For a study that tests how well sensors work, the sensor nodes should be placed at several places where traffic is common and run for a long enough time to cover different types of traffic and weather.Each data entry needs to include the time it was recorded, the location it comes from, the levels of pollutants, temperature, humidity, and any other available information about weather or traffic.If possible, data from a properly checked device or a known monitoring station should also be gathered to check if the sensors are accurate.The exact places where the sensors were put, how long the data was collected, the types of sensors used, how often data was taken, and the total amount of data should all be clearly written down in the final report.

B. Feature Engineering and Targets

Useful features include the latest pollutant levels, average values over time, previous readings, the time of day, the day of the week, temperature, humidity, rainfall, wind speed, and a measure that shows traffic volume.When predicting for the next hour, the goal is to forecast the pollutant level or an air quality index-related value that will occur one hour after the data window.For classification, the category should be based on a clear and suitable air quality index standard.The set of features used must only include information that would be available when making the prediction.

C. Training and Validation

Data should be split in order over time into three parts: training, validation, and a separate test period.Steps like filling in missing data, adjusting values to a common scale, choosing important features, and fine-tuning the model are done using the training data and then used on the later periods.The baselines that are compared should include the persistence method, which uses the most recent data point as the prediction, a basic regression model, and at least one model based on trees.Possible advanced models to consider could be random forest regression, NARX or LSTM networks, and SV-CNN for classification.The best settings for the models, called hyperparameters, should be chosen using the validation data, not the final test data.

D. Evaluation Metrics

For numerical predictions, include the MAE, RMSE, and R^2 values along with the forecast time period and units.In classification tasks, report accuracy, precision, recall, and F1 score, ideally including results for each class and a confusion matrix.Since high-pollution events are particularly significant, focus on the recall for the high-risk category and the count of false alarms.The operational evaluation should also note inference time, success rate of data transmission, rate of missing data, and usage of resources on edge devices.Remember that regression and classification metrics do not measure the same aspects and should not be treated as equivalent..

V. RESULTS AND DISCUSSION

This paper presents a proposed framework and does not include any new experiments or newly collected traffic data.As a result, no performance metrics are claimed for the integrated system.The results from the referenced studies support the design decisions made, but the specific numbers cannot be directly compared because the datasets, target variables, prediction timeframes, hardware, and evaluation methods vary across the studies.These sources offer valuable insights.Banciu et al.shared their best random-forest regression error results for their specific AQI targets [3].Moursi et al.found that NARX/LSTM works well for cloud forecasting, while NARX/XGBRF provides a good balance of speed and accuracy on a Raspberry Pi 4 [2].Kumar and Doss introduced a portable, cloud-connected monitoring concept that includes a deep-learning forecasting model [1].Alnowaiser et al.reported strong classification results using SV-CNN with KNN imputation [4]. These results suggest that testing a specific mix of regression, time prediction, handling missing data, and edge computing could be effective.Before using this approach in real-world settings, it should be tested in areas with heavy traffic.When comparing different models, it's important to use the same time period for

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

testing and define the goals in the same way for all models.The results should be shared separately based on location and time of year if the data allows.Factors like sensor wear, changes in weather, road work, and shifts in traffic flow might affect how well the model works after it's deployed.Because of this, the model may need regular checks and updates to keep it accurate

DESIGN OPTIONS SUGGESTED BY THE SOURCE PAPERS

System need Candidate approach Source basis
Portable distributed sensing Location-tagged IoT nodes and cloud upload [1]
Short-term PM2.5 forecast NARX with LSTM or tree-based model [2]
Numeric AQI prediction Random-forest regression / neural network baseline [3]
Missing-value handling and classification Evaluate KNN imputation and SV-CNN ensemble [4]
Low-connectivity response Edge inference plus cloud synchronization [2]

VI. LIMITATIONS AND FUTURE WORK

The design has some issues that need to be fixed when it's built.Cheap sensors can become less accurate over time and might be influenced by factors like humidity, temperature, where they're placed, and sensitivity to other things.Information about traffic and weather might not be available in every location.Using KNN to fill in missing data can be slow for big sets of information and might hide sudden pollution increases if not used carefully.Deep learning models that look at time patterns might need more past data and powerful computers than small sensors can provide.Also, doing well on one set of data doesn't mean it will work well in another city or during a different time of year.Future work should create a dataset from several places close to busy roads, check the sensor readings against high-quality instruments, and test the system under different traffic and weather situations.Other research could compare using the system in the cloud versus on local devices, measure how much power it uses and how fast it works, and test moving knowledge from one place to another.

A public dashboard should display details about uncertainty or confidence when they are available, use clear signs to show how reliable the data is, and offer a method for users to point out any problems.

VII. CONCLUSION

This paper introduces an IoT-based smart system designed to predict air quality in busy traffic areas.It uses a combination of distributed sensors, a microcontroller, network communication, cloud storage, data processing, machine learning models, and a dashboard with alert features.Four studies were presented, each offering supporting evidence in areas like portable monitoring, cloud-based AQI regression, PM2.5 forecasting on edge devices, and classification using KNN imputation.The framework brings together these elements while keeping regression and classification tasks separate and avoids making unsupported claims about new accuracy results.To prove how well it works, it needs to be tested in a controlled setting using local traffic-area data, clear comparison standards, suitable performance measures, and confirmed sensor reliability..

REFERENCES

[1] T. Kumar and A. Doss, “AIRO: Development of an Intelligent IoT-based Air Quality Monitoring Solution for Urban Areas,” Procedia Computer Science, vol. 218, pp. 262–273, 2023, doi: 10.1016/j.procs.2023.01.008.
[2] A. S. Moursi, N. El-Fishawy, S. Djahel, and M. A. Shouman, “An IoT enabled system for enhanced air quality monitoring and prediction on the edge,” Complex & Intelligent Systems, vol. 7, pp. 2923–2947, 2021, doi: 10.1007/s40747-021-00476-w.
[3] C. Banciu, A. Florea, and R. Bogdan, “Monitoring and Predicting Air Quality with IoT Devices,” Processes, vol. 12, no. 9, Art. no. 1961, 2024, doi: 10.3390/pr12091961.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

[4] K. Alnowaiser, A. A. Alarfaj, E. A. Alabdulqader, M. Umer, L. Cascone, and B. Alankar, "IoT based smart framework to predict air quality in congested traffic areas using SV-CNN ensemble and KNN imputation model," Computers and Electrical Engineering, vol. 118, Art. no. 109311, 2024, doi: 10.1016/j.compeleceng.2024.109311