ISSN 2007-9737

IoT-Based Intelligent System for Monitoring and Predicting Urban Air Pollution: Spatial and Temporal Analysis in Seoul

Aaron Mina-Rios[1], Sebastian Albornoz-Yanac[1], Sebastian Lucich-Muñoz[1], Javier Gamboa-Cruzado[2], Alex Salazar-Marzal[2], Carlos Chávez-Herrera[2], Carlos Del-Valle-Jurado[2], Obdulia Pichardo-Lagunas[3,*]

[1] Universidad Nacional Federico Villarreal, Peru

[2] Universidad Nacional Mayor de San Marcos, Peru

[3] Instituto Politécnico Nacional, UPIITA, Mexico

{2021018505, 2021018479, 2022016372}@unfv.edu.pe, {jgamboac, alexmelecio.salazar, cchavezh, cdelvallej}@unmsm.edu.pe, opichardola@ipn.mx

Abstract. Urban air pollution constitutes a critical challenge for public health and environmental sustainability, while traditional monitoring approaches present limitations in coverage, timeliness, and data quality. In this context, the objective of this research was to develop and implement an IoT-based intelligent system for monitoring and analyzing air pollution in the city of Seoul, integrating data acquisition, cloud storage, analytical visualization, and experimental evaluation of its performance. Methodologically, an applied approach with a pure experimental design was adopted, and the Scrum methodology was employed for the development of the solution. The analysis was conducted using the public dataset Air Pollution in Seoul, and a control group with manual processes was compared with an experimental group using automated processing. The results show a consistent classification performance of the system, with values of Accuracy = 0.822, Precision = 0.868, Recall = 0.822, and F1-Score = 0.836, as well as a substantial reduction in report preparation time, in the percentage of incorrect data, and in missing data. Likewise, the spatial and temporal analysis of PM2.5 made it possible to identify persistent hotspots and relevant variation patterns at the urban scale. In conclusion, the proposed system improves operational efficiency, data reliability, and the analytical capacity of environmental monitoring, representing a robust alternative to support decision-making in smart urban environments.

Keywords*.* Internet of things, IoT, environmental monitoring, air pollution, air quality, dataset, sensors.

1 Introduction

The degradation of air quality represents a critical challenge for public health and environmental sustainability, particularly in urban environments where increased vehicular traffic, industrial activity, and population density intensify the emission of atmospheric pollutants. However, traditional monitoring methods present significant limitations due to their dependence on fixed monitoring stations with limited spatial coverage and low measurement frequency, which makes it difficult to capture the temporal and spatial variability of pollution. In this context, the need arises to develop intelligent systems capable of integrating distributed sensors, continuous connectivity, and automated data analysis in order to improve environmental monitoring and support decision making based on timely and reliable information. First, several studies have shown that the integration of IoT with machine learning techniques strengthens air quality monitoring and prediction by enabling the identification of pollution patterns, the classification of risk levels, and the anticipation of critical episodes with greater accuracy in dynamic urban environments [1,19,36]. Likewise, various studies agree that distributed networks of low-cost IoT sensors expand the spatial coverage of environmental monitoring and enable hyperlocal real time analysis, which is particularly valuable for the early detection of pollution events and risk based management in smart cities [2,3,39]. In addition, the authors in [4,28,40], in their research on intelligent systems applied to environmental control, demonstrate that the articulation of distributed sensors, cloud processing, artificial intelligence, and connected platforms improves the predictive capacity of the system and strengthens proactive decision making in response to atmospheric pollution.

Along the same line, it has been reported that open, scalable, and economically accessible solutions represent a promising pathway for democratizing environmental monitoring. In particular, the works in [5,6,38] show that low cost sensors and open architectures can generate useful data for continuous surveillance, although their consolidation depends on rigorous field validation and calibration processes. On the other hand, some studies have emphasized the importance of open platforms, data visualization, and user interaction as essential components of the IoT ecosystem. In this regard, the contributions in [7,9,29] suggest that monitoring systems should not only capture data but also facilitate their access, interpretation, and use through interfaces, open servers, and applications oriented to different contexts of use. Furthermore, recent literature has also highlighted applications focused on health and the monitoring of indoor or institutional environments. Studies [8,24,35] reveal that IoT architectures allow the characterization of exposure conditions in hospitals, buildings, and other sensitive spaces, strengthening the protection of vulnerable populations and supporting corrective actions based on real time environmental data.

From another perspective, the expansion of these infrastructures has revealed that monitoring reliability does not depend solely on data acquisition but also on its security, traceability, and protection. In this framework, the authors in [10,13,41] emphasize that IoT systems generate large volumes of information and therefore require robust mechanisms of privacy, authentication, blockchain, and data protection to ensure trustworthy use in connected urban environments. Similarly, several studies have demonstrated that dense sensor networks and the integration of multiple information sources significantly improve the spatial resolution of atmospheric monitoring. In particular, [12,15,22] show that the combination of distributed sensors, satellite data, and analytical models makes it possible to characterize urban pollution more accurately, especially for critical variables such as PM2.5 and PM10. Regarding the availability of empirical evidence, there is an increasing trend toward the construction of open datasets and spatiotemporal repositories that strengthen the training and validation of intelligent models. Studies [14,27,32] show that datasets obtained through IoT sensors, crowdsensing, or urban platforms are essential for developing more robust predictive models, although limitations related to coverage, quality, and heterogeneity of collected variables still persist.

In addition, other studies have explored the monitoring of pollutants in specific or highly complex operational scenarios, such as refineries, subway tunnels, or open pit mines. The results reported in [16,17,37] confirm that IoT solutions allow the capture of valuable information in these environments, although their performance still depends on factors such as limited coverage, extreme exposure of equipment, or the need to adapt sensors to particular operational conditions. On the other hand, an important group of studies has addressed the problem of data reliability through intelligent correction, fault tolerance, and dynamic sensor calibration. In this regard, [20,21,23] show that incorporating automatic mechanisms for fault detection, quality control, and collaborative calibration improves system accuracy and reduces the uncertainty associated with the use of low cost sensors in urban networks. In addition, the literature has advanced toward more flexible, mobile, and distributed monitoring schemes by combining drones, edge computing, and local processing nodes. The works in [25,30,34] demonstrate that these strategies increase processing speed, expand coverage in complex areas, and strengthen decision support services, although improvements are still required in stability, energy consumption, and field accuracy.

In continuity with this perspective, several authors have proposed comprehensive intelligent frameworks for environmental assessment by combining IoT, advanced analytics, atmospheric modeling, and calibrated sensor networks. Studies [26,31,33] agree that these approaches increase the usefulness of monitoring for public management and comparative analysis between cities, although they still require greater methodological standardization and validation in heterogeneous urban contexts. Finally, another line of research has begun to link atmospheric monitoring with human exposure and well being by integrating high resolution environmental data with biometric, multisource, and hyperlocal exposure variables. In this sense, [18,22,39] suggest that the future of intelligent monitoring lies not only in measuring pollutants with greater precision but also in translating this information into useful knowledge to understand real health risks and guide more targeted urban interventions. Overall, the reviewed literature demonstrates substantial advances in IoT sensors, intelligent platforms, calibration, prediction, and environmental analytics. However, challenges remain related to the accuracy of low cost sensors, the integration of multiple sources, methodological standardization, data security, and validation in complex urban scenarios [10,21,26]. These gaps justify the development of IoT-based intelligent systems oriented not only toward continuous monitoring but also toward the reliable prediction of urban air pollution through robust, scalable, and analytically consistent architectures.

Despite the advances reported in IoT sensor networks, analytical platforms, and predictive models for air quality monitoring, the literature reveals persistent limitations related to the operational integration of these technologies into intelligent architectures capable of simultaneously managing the capture, processing, storage, and visualization of environmental data. Likewise, many studies focus on specific components such as sensors, predictive models, or datasets without comprehensively addressing the full cycle of data acquisition, processing, and analysis within a unified intelligent system. This technological fragmentation limits the implementation of robust solutions aimed at continuous monitoring and reliable prediction of air pollution in complex urban environments. In response to these limitations, it becomes necessary to develop solutions that integrate environmental sensors, data transmission, cloud infrastructure, and advanced analytical tools within a coherent IoT-based intelligent system architecture. This approach improves continuous monitoring capacity, optimizes the management of large volumes of environmental data, and facilitates the generation of useful information for decision making. In particular, the development of systems that integrate data acquisition, automated processing, and visualization through interactive dashboards contributes to strengthening urban environmental surveillance and supporting evidence based management strategies. The objective of this research is to develop and implement an IoT-based intelligent system for monitoring and analyzing urban air pollution in the city of Seoul, integrating data processing, cloud storage, and visualization through analytical dashboards. The system aims to improve environmental information management and provide tools that facilitate the interpretation of pollution trends and support data driven decision making.

The paper is organized as follows. Section 2 presents the Background, where the fundamental concepts related to IoT, environmental sensors, and air quality are reviewed. Section 3 describes the Research Method, detailing the approach, the data collection techniques, the Scrum methodology applied for the planning and iterative development of the system, as well as the procedure followed to implement and evaluate the monitoring process. Section 4 corresponds to the Case Study, where the architecture of the proposed IoT system, the components used, and the real context in which it was applied are explained. Section 5 presents the Results and Discussion, including performance analysis, interpretation of environmental data, and comparison with previous studies. Section 6 presents the Conclusions, highlighting the contributions, limitations, and opportunities for future improvement.

2 Background

2.1 Internet of Things (IoT)

The Internet of Things (IoT) is understood as a technological ecosystem composed of physical devices capable of capturing information from the surrounding environment and transmitting it through communication networks.

According to Banciu [4], IoT can be conceived as “a distributed architecture based on sensors that enables the collection of real time information for its analysis and to support decision making processes.”

In the same line, Fikri [9] states that IoT integrates sensors, communication modules, and cloud platforms, which makes possible the continuous monitoring of environmental variables and the automated processing of the collected data.

Likewise, Collado [6] highlights that this technology facilitates the development of low cost monitoring infrastructures capable of covering extensive areas through the use of accessible hardware and open protocols.

Finally, Daffa Prebian [7] emphasizes that IoT systems strengthen environmental monitoring and control capabilities by generating constant data streams, allowing atmospheric conditions to be evaluated with greater precision and at more appropriate temporal intervals.

2.2 Environmental Pollution

Atmospheric pollution is defined as the accumulation of particles and gases in the air that generate adverse effects on both human health and ecosystem balance. Abdelmalek [1] notes that air quality is usually evaluated through indices that integrate concentrations of toxic gases such as NO_2, CO, and SO_2 together with suspended particulate matter, which allows a comprehensive estimation of their impact on public health. Likewise, Alsamrai [3] emphasizes that atmospheric pollution presents a dynamic behavior, being able to vary within minutes due to factors such as traffic intensity, industrial activities, and meteorological conditions.

In this context, Jang [15] indicates that among the most relevant pollutants are fine particulate matter PM2.5 and PM10, whose high concentrations are associated with a significant increase in the risk of respiratory and cardiovascular diseases. Finally, Karnati [19] highlights that the incorporation of IoT sensors and data analysis techniques makes it possible to identify pollution patterns, analyze trends, and anticipate critical episodes, providing key information for environmental management in urban environments.

2.3 IoT-Based Intelligent System and Urban Air Pollution Monitoring

IoT-based intelligent systems have transformed traditional approaches to monitoring and predicting air pollution in urban environments. IoT is configured as a distributed architecture that integrates sensors, communication networks, and processing platforms capable of collecting and transmitting environmental data in real time, expanding spatial and temporal coverage compared with conventional systems based on fixed monitoring stations.

Through environmental sensors and communication technologies, IoT systems enable the capture of concentrations of pollutants such as PM2.5, PM10, NO_2, CO, SO_2, and O_3, generating continuous streams of information that can be analyzed using data analysis techniques and time series models.

This technological integration facilitates the identification of pollution patterns and the development of predictive models that contribute to anticipating critical episodes of air quality deterioration and strengthening environmental management in smart cities.

3 Research Method

The research adopts an applied approach aimed at the development and evaluation of an IoT-based intelligent system for monitoring urban air pollution.

To achieve this, a software development methodology and an experimental design are integrated to analyze the system’s performance in terms of efficiency and data quality.

The methodological process includes the application of the Scrum methodology for system development, the operationalization of variables related to environmental monitoring, and the use of the public dataset Air Pollution in Seoul for the analysis of atmospheric pollutants in the city of Seoul.

ISSN 2007-9737

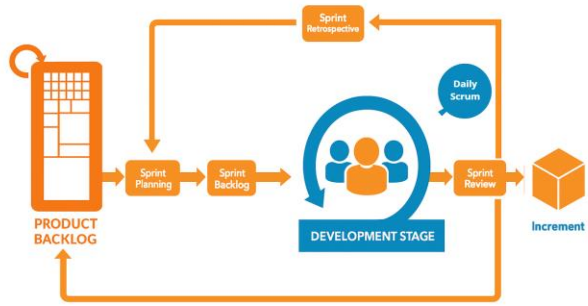


Fig. 1. Scrum stages

Table 1. Operationalization of the dependent variable (Environmental pollution)

Indicator	Index	Unit of Measure-ment	Unit of Observation
Report preparation time	[0 - 7]	min	OPC-R1 Sensor / Air Pollution in Seoul
% of incorrect data	[0 - 5]	%	OPC-R1 Sensor / Air Pollution in Seoul
% of missing data	[0 – 5]	%	OPC-R1 Sensor / Air Pollution in Seoul

3.1 Solution Development Methodology: Scrum Methodology

In this research, the Scrum methodology was used for the development of the IoT-based intelligent system designed for monitoring air pollution.

According to Gamboa-Cruzado and colleagues [11], Scrum is an agile methodology oriented toward collaborative work and continuous improvement, particularly suitable for software development projects. This methodology is structured around clearly defined roles and organizes the development process into iterative cycles known as Sprints, which establish specific periods and concrete objectives for the progressive delivery of system functionalities.

This approach allows progress to be evaluated periodically, facilitating adaptation and improvement of the product during its development, as shown in Figure 1.

The stages of the Scrum process are as follows:

– **Sprint Planning**: In this phase, the Product Owner presents the prioritized elements of the product backlog, and together with the development team defines the tasks that will be addressed during the next Sprint.
– **Sprint**: This is the period in which the team develops the selected activities while conducting the Daily Scrum to review progress, identify obstacles, and adjust the work.
– **Sprint Review**: At the end of the Sprint, the team presents the developed increment and receives feedback from the Product Owner and stakeholders to guide future improvements.
– **Sprint Retrospective**: In this stage, the team reflects on the performance of the Sprint, identifying achievements, difficulties, and opportunities for improvement that allow the optimization of the development process in subsequent cycles.

3.2 Applied Research Methodology

In order to evaluate the impact of the IoT-based intelligent system on environmental pollution monitoring, an applied research methodology was adopted.

This approach makes it possible to empirically analyze the system's performance through specific indicators that measure operational efficiency and the quality of the generated data.

3.2.1 Operationalization of variables

The operationalization of variables enables the transformation of theoretical concepts into observable and measurable indicators that facilitate empirical evaluation. In this study, the dependent variable environmental pollution is operationalized through indicators that allow the assessment of the efficiency of the IoT-based intelligent system and the quality of the data generated by the monitoring sensors, as presented in Table 1.

The indicators presented allow a comprehensive evaluation of both the data quality and the operational efficiency of the IoT-based intelligent system for environmental pollution monitoring. The report preparation time indicator measures the system's capacity to process and deliver information from the moment data are loaded into the database until they are visualized in the generated reports. Likewise, the percentage of incorrect data and the percentage of missing data constitute essential metrics for determining the reliability and completeness of pollutant records.

Both indicators are evaluated considering a maximum threshold of 5%, which ensures a minimum reliability level of 95% in valid system records. Finally, user satisfaction provides a qualitative dimension that allows the evaluation of system acceptance, usability, and user perception in the work environment.

3.2.2 Research Design

The research corresponds to an applied study with a pure experimental design, in which the results obtained between an experimental group and a control group are compared in order to evaluate the impact of the IoT-based intelligent system on the efficiency of data processing and the quality of the generated information:

$$RG_e \quad X \quad O_1, \tag{1}$$

$$RG_e \quad \cdots \quad O_2. \tag{2}$$

Data were collected from the experimental group (Ge), to which the independent variable corresponding to the IoT-based intelligent system was applied, and from the control group (Gc), in which this intervention was not implemented.

The sample elements, equivalent to 5% of the total records, were selected randomly. The technological stimulus, namely the IoT-based system, was applied to this sample.

For the measurement of the report preparation time indicator, in the control group (Gc) timed measurements were conducted to determine the time required to manually update the data in the reports. In the experimental group (Ge), the delay time between the loading of data into the database and their automatic update in the report generated by the system was measured.

Regarding the indicators percentage of missing data and percentage of incorrect data, in the control group (Gc) a manual analysis of the sample records was performed using Excel spreadsheets. In the experimental group (Ge), two automated analysis routines implemented in Python were used to evaluate the quality and completeness of the data contained in the dataset.

Finally, to measure user satisfaction, surveys were applied in both groups. In the control group (Gc), the interaction with the initial Excel-based system was evaluated, while in the experimental group (Ge) the perception of the reports generated by the IoT-based intelligent system was analyzed.

3.2.3 Population and Sample

The study population consisted of all measurement records corresponding to six atmospheric pollutants (SO_2, NO_2, O_3, CO, PM_{10}, and $PM_{2.5}$) collected in the city of Seoul during the period 2017–2019, with a total of N=647,511 observations.

For the experimental analysis, a sample equivalent to 5% of the population was selected for the execution of the tests and evaluations of the proposed system, obtaining a total of n = 32,376 records.

3.2.4 Data Collection Procedure

In this research, the data collection instrument consisted of the selection and analysis of a public dataset, specifically the dataset "Air Pollution in Seoul", available on the Kaggle platform.

The techniques employed included the indirect observation of the records and database consultation to obtain historical information related to air quality in the city of Seoul.

4 Case Study

In the development of the IoT system for environmental pollution monitoring in the city of Seoul, the Scrum methodology was applied, structuring the process into four main stages oriented toward the design, implementation, and evaluation of the system.

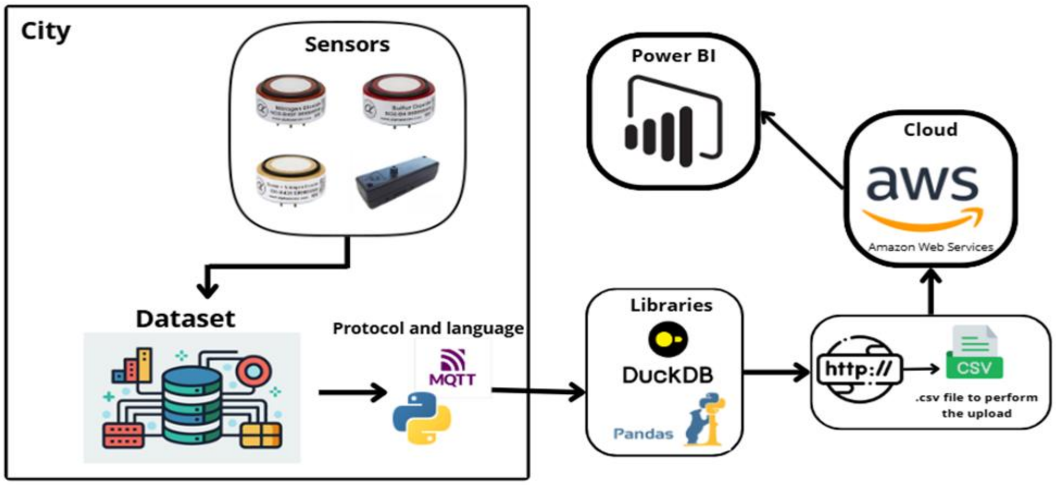


Fig. 2. Architecture of the IoT-Based Intelligent System

Table 2. Functional Requirements

Id	Requirements
RF01	The system must manage information on the following atmospheric pollutants: PM2.5, PM10, NO_2, SO_2, CO, and O_3.
RF02	The system must store the measurements in a cloud database for consultation and analysis.
RF03	The system must display the information on a dashboard containing charts, maps, and reports.
RF04	The system must allow the creation of reports on air quality based on filters for territory and time periods.
RF05	The system must allow data export in CSV format and must be compatible with analysis tools such as Power BI.
RF06	The system must support at least 3 years of historical data without losing information.
RF07	The system must allow comparison of pollution levels between different zones and districts of the city of Seoul.
RF08	The system must allow filtering of information by pollutants, districts or zones, and years.

4.1 Sprint Planning

At this stage, the overall architecture of the system was defined and the functional requirements that guide its development were identified. In addition, the work plan was organized into four Sprints, establishing specific objectives and deliverables for each iteration of the development process.

4.2 System Architecture

The system architecture describes the technological infrastructure responsible for the capture, processing, and storage of data generated by environmental sensors installed in the city.

This workflow integrates IoT devices, communication protocols, and cloud services to ensure efficient information management, as shown in Figure 2.

This approach is consistent with Tupac-Agüero [43] in recent research, where it is highlighted that the integration of sensor platforms and IoT technologies facilitates the real time collection of critical variables and strengthens intelligent environmental monitoring systems oriented toward pollutant detection.

ISSN 2007-9737

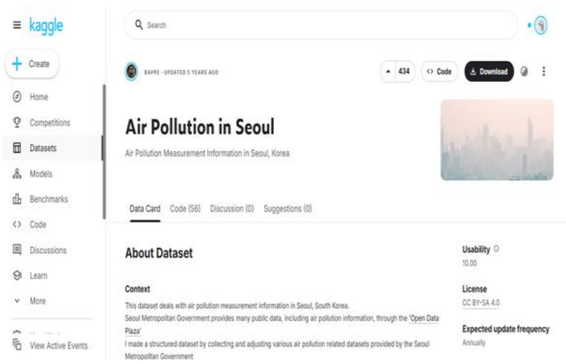


Fig. 3. Selected dataset (Kaggle)

4.3 Functional Requirements

The identified data made it possible to define the functional requirements of the IoT-based intelligent system, which establish the essential functions for its operation. These include the management and storage of information on atmospheric pollutants, the visualization of data through dashboards, the generation and export of reports, as well as the comparison and filtering of records across different zones and time periods, allowing a more efficient and comprehensive environmental analysis.

Finally, for the organization of the Sprints, the Product Owner prioritized the elements of the product backlog, including the selection of the dataset, the cleaning and preparation of the data, the transmission of information to the cloud environment, and the development of the dashboard. Based on this prioritization, the team defined the scope and objectives of the following Sprints:

- **Sprint 1**: Project definition and dataset selection
- **Sprint 2**: Cleaning and loading of the selected dataset
- **Sprint 3**: Data transmission and storage
- **Sprint 4**: Development of interactive dashboards in Power BI.

4.4 Sprint

During this phase, the Sprints defined in the planning stage were executed, each oriented toward the development of functional increments of the IoT system for monitoring environmental pollution in Seoul. The process was organized into four consecutive Sprints, each with specific objectives and clearly defined deliverables.

4.4.1 Sprint 1: Project Definition and Dataset Selection

In this first Sprint, the fundamental elements of the project were defined, including the identification of a reliable data source, the delimitation of the system scope, and the selection of the application domain that would guide the development of the subsequent Sprints. This stage made it possible to establish the overall approach of the project before initiating the technical implementation phases.

The elements defined in this phase were the following:

- **Selected sector:** environmental pollution
- **Project scope:** development of an IoT system for the capture, processing, and visualization of environmental data.
- **Selected dataset:** *Air Pollution in Seoul*, corresponding to the period 2017–2019, with a total of 647,511 records, used as the main database for development, as shown in Figure 3.

4.4.2 Sprint 2: Cleaning and Loading of the Selected Dataset

In this Sprint, the cleaning and preparation process of the selected dataset was carried out. During this stage, values of 0 and -1 were identified in the columns corresponding to the six pollutants (SO_2, NO_2, O_3, CO, PM_{10}, and PM2.5), which were replaced with null values due to their invalid nature for analysis.

Subsequently, the rows containing null values across all pollutant variables were removed. The remaining NaN values were imputed using the median of each variable in order to preserve the distribution of the data. In addition, records that exceeded the maximum permissible value for each pollutant were adjusted to the corresponding upper limit.

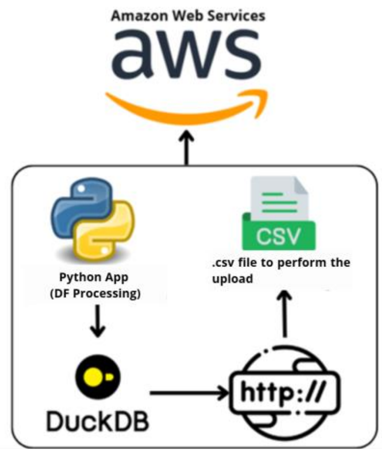
For these operations, data analysis libraries such as Pandas and DuckDB were used, which allowed the dataset to be standardized and its

ISSN 2007-9737


```python
# ====== 4. Data cleaning ======
cols = ["so2", "no2", "o3", "co", "pm10", "pm2_5"]

# --- 4.1. Replace 0 and -1 with NaN
df[cols] = df[cols].replace({-1: np.nan, 0: np.nan})

# --- 4.2. Delete rows where all 6 contaminants are NaN
filas_iniciales = len(df)
df = df.dropna(subset=cols, how="all")
filas_finales = len(df)
print(f"🧹 Rows removed because they all contain NaN contaminants: {filas_iniciales - filas_finales}")

# --- 4.3. Outlier rules (clip to maximum allowed)
maximos = {"so2": 1, "no2": 2, "co": 50, "o3": 0.5, "pm10": 600, "pm2_5": 500}
for col, max_val in maximos.items():
    df[col] = np.where(df[col] > max_val, max_val, df[col])

# --- 4.4. Calculate the median of each column (ignoring NaN)
medianas = df[cols].median()

# --- 4.5. Fill NaN values with the median of their respective column
df[cols] = df[cols].fillna(medianas)

# --- 4.6. Show applied medians
print("\n📊 Mediums used for replacement:")
for col, mediana in medianas.items():
    print(f" - {col}: {mediana}")
```

```
PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   PORTS

(entornoiot) PS C:\Users\sebas\OneDrive\Desktop\ProyIoT> & C:/Users/sebas/OneDrive/Desktop/ProyIoT/entornoiot/Scripts/python.exe c:/Users/sebas/OneDrive/Desktop/ProyIoT/Carga_CSV2.py
🧹 Rows removed because they all contain NaN contaminants: 6119

📊 Mediums used for replacement:
 - so2: 0.004
 - no2: 0.025
 - o3: 0.021
 - co: 0.5
 - pm10: 35.0
 - pm2_5: 19.0

✅ Clean CSV file uploaded to S3: s3://proyectiot
(entornoiot) PS C:\Users\sebas\OneDrive\Desktop\ProyIoT>

Ln 64, Col 32   Spaces: 4   UTF-8   CRLF   {} Python   3.12.6 (entornoiot)
```

Fig. 4. Python data cleaning code and its execution




```python
# ====== 5. Connect to DuckDB with AWS S3 support ======
con = duckdb.connect()
con.execute("""
INSTALL httpfs;
LOAD httpfs;
""")

# ====== 6. Register the cleaned DataFrame ======
con.register("df_view", df)

# ====== 7. Export to S3 as CSV ======
con.execute("""
COPY (SELECT * FROM df_view)
TO 's3://proyectiot/IoTContaminacion.csv'
WITH (FORMAT CSV, HEADER, OVERWRITE_OR_IGNORE);
""")

print("\n✅ Clean CSV file uploaded to S3: s3://proyectiot")
```

Fig. 5. Data transmission via HTTP and connection to AWS S3

consistency to be ensured before its integration into the system, as shown in Figure 4.

4.4.3 Sprint 3: Data Transmission and Storage

Once the data were processed and cleaned, the transmission logic was implemented using the HTTP protocol, simulating the sending of measurements from IoT sensors. This mechanism made it possible to reproduce the typical communication flow between environmental data capture devices and a cloud storage infrastructure.

Subsequently, the data were stored in Amazon Web Services (AWS S3) by configuring the buckets, access policies, and the necessary permissions to ensure the security and availability of the information. In addition, stability tests were performed on the transmission and upload of the CSV files to the cloud environment. This process made it possible to validate the system's capability to receive, transmit, and store environmental data securely and continuously, as shown in Figures 5.

4.4.4 Sprint 4: Development of interactive dashboards in Power BI

This Sprint focused on the development of interactive dashboards for the visualization of air quality, designed for both desktop and mobile

ISSN 2007-9737

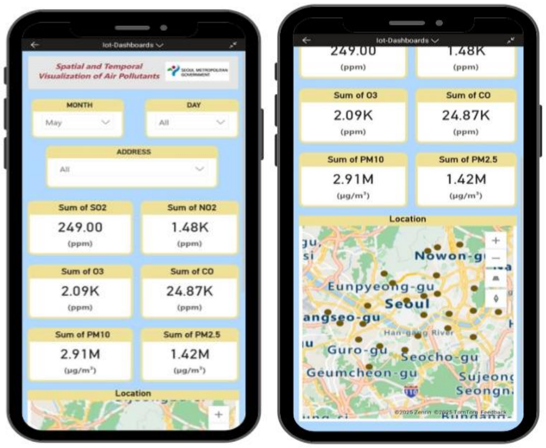


Fig. 6 and **Fig. 7.** Temporal (Fig. 6) and spatial (Fig. 7) visualization of atmospheric pollutants – Mobile

environments. For this purpose, Power BI was used, enabling the creation of dynamic visualizations that facilitate the exploration and analysis of environmental data.

The dashboards incorporate different types of visualizations, including line charts, donut charts, matrices, and geospatial maps, as well as interactive filters that allow the information to be analyzed by year, month, and district. In addition, key pollution indicators were integrated to interpret trends and variations in pollutant concentrations across different areas of Seoul.

These visualizations make it possible to identify spatial and temporal patterns of pollution, facilitating data interpretation and supporting decision making in environmental management, as shown in Figures 6 and 7.

4.5 Sprint Review

During the Sprint Review, the progress achieved in each stage of the project was presented, including the definition of the project scope, the cleaning and preparation of the dataset, as well as the transmission and storage of data in the cloud through AWS S3.

In addition, the transformation of the dataset into an entity–relationship model was presented, which allowed the information to be properly structured through entities such as districts, stations, pollutants, and measurements, facilitating its organization and subsequent analysis. Finally, the dashboards developed in Power BI were evaluated to verify that they correctly represented the processed information and complied with the defined functional requirements, as shown in Figures 8 and 9.

4.6 Sprint Retrospective

During the retrospective, the team evaluated the development of the four Sprints and the activities carried out at each stage of the project. The main achievement identified was the appropriate organization of the work process, from the initial project definition to the implementation of the final dashboards. In addition, the logical sequence of the activities developed was highlighted, which included dataset selection, data cleaning and preparation, information transmission, and the development of the visualizations.

As aspects for improvement, the need to strengthen the documentation of some technical processes and to optimize communication during the integration of the results from each Sprint was identified. Finally, the team agreed to maintain structured planning and conduct periodic reviews of project progress in order to improve efficiency in future developments.

5 Results and Discussion

This section presents the results obtained from the experimental evaluation of the IoT-based intelligent system for monitoring and predicting air pollution in the city of Seoul.

First, the performance metrics of the classification model applied to the environmental dataset are analyzed in order to evaluate the system's capacity to identify and categorize pollution levels.

Subsequently, spatial and temporal pollution patterns are examined through geospatial analysis and PM2.5 time series.

Finally, the results obtained between the control group and the experimental group are compared through operational system indicators, allowing a discussion of the improvements achieved in terms of efficiency, data quality, and reliability.

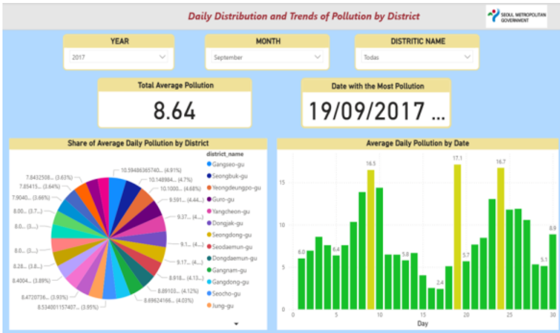


Fig. 8. and **Fig. 9.** Daily distribution and pollution trends by district – Desktop (Fig. 8) and Mobile (Fig.9)

📈 IoT System Performance Evaluation:

✅ Accuracy: 0.822
🎯 Precision: 0.868
🔎 Recall: 0.822
⚖ F1-Score: 0.836

Fig. 10. Performance metric results of the IoT system

5.1 Classification Performance of the IoT System

To evaluate the performance of the IoT-based intelligent system, a classification model using machine learning techniques was applied to the environmental dataset. The model was trained using variables related to atmospheric pollutants recorded in the dataset, and its performance was evaluated using standard classification metrics such as Accuracy, Precision, Recall, and F1-score. These metrics make it possible to measure the model's ability to correctly identify pollution levels and classify environmental records according to their characteristics.

- **Accuracy**: quantifies the overall proportion of correct predictions:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}. \tag{3}$$

- **Precision:** measures the accuracy in identifying positive instances:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

- **Recall:** indicates the system's ability to detect all positive cases:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

- **F1-score:** combines Precision and Recall through their harmonic mean:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

Next, the results obtained for each of the evaluated metrics are presented, allowing a clear visualization of the performance of the IoT system, as shown in Figure 10.

The results obtained show that the IoT system presents consistent performance in the classification of air pollution levels. The evaluation metrics reveal an appropriate balance between precision, sensitivity, and predictive stability, suggesting that the model is capable of identifying relevant patterns in environmental data and adequately discriminating among different pollution categories.

As observed in the results, Precision is the highest metric, suggesting that the system produces relatively few false positives when identifying pollution episodes or categories. This behavior can be explained by the model's ability to recognize sufficiently defined patterns in the input data, likely favored by the structure of the dataset and by the presence of stable relationships between environmental variables and pollution levels. The results also show that the fact that Accuracy and Recall present the same value indicates that, although the system correctly classifies a significant proportion of cases, it still faces difficulties in capturing some real events. This situation may be associated with overlap between categories, the natural variability of pollutants, or subtle differences between observations with similar behavior. Based on the observed distribution of the metrics, the F1-score confirms a balanced performance between precision and sensitivity, suggesting that the model is not excessively biased toward a single performance dimension. From a causal perspective, this pattern may be explained by the fact that the system adequately discriminates the most representative cases but loses effectiveness in borderline situations or in classes with greater temporal and spatial heterogeneity.

When comparing the system's performance with previous studies, a consistent trend with the literature on IoT-based environmental monitoring and machine learning can be observed. Indeed, several studies report high levels of performance in classification metrics. For example, Alsamrai and colleagues [3] report values close to 99% using models such as Gradient Boosting, Random Forest, and SVC, while Yildiz and Sucuoglu [36] achieve approximately 91% in AQI category classification using IoT-based forecasting architectures. Similarly, Islam and collaborators [14] show that algorithms such as KNN, Naive Bayes, and Random Forest reach values between 93% and 97.2% in Accuracy, Precision, Recall, and F1-score, and Khan and co-authors [20] demonstrate that hybrid models such as LSTM-KNN and CNN-LSTM achieve performance levels close to 96–97% in pollutant prediction. Likewise, Parra-Medina and their research team [25] report values close to 94% in IoT systems using calibrated sensors and LoRa communication, while Zafra-Pérez and other researchers [37] show F1 values close to 95% in demanding environments. In addition, Karnati [19] highlights that the integration of sensors with IoT platforms increases system sensitivity due to noise reduction and the continuous availability of real-time data. Although the results of the present study are slightly lower than some of these previous works, these differences can be explained by the use of operational conditions closer to real scenarios, where meteorological variability, heterogeneity of emission sources, environmental noise, and the limitations of low-cost IoT sensors introduce greater analytical complexity. In contrast, several previous studies employ highly curated datasets, specifically optimized architectures, or more controlled experimental environments, which tend to produce higher metrics. Nevertheless, the values obtained remain within a comparable trend and reflect an appropriate balance between predictive accuracy and operational feasibility. Therefore, they support the usefulness of the system for real urban monitoring applications. Consequently, the performance achieved can be considered consistent with the state of the art in IoT systems for air quality monitoring and confirms that the integration of connected sensors and machine learning models constitutes a valid strategy for the classification of atmospheric pollutants. Overall, the results are not only comparable with recent literature but also demonstrate the feasibility of implementing robust solutions in dynamic environmental contexts.

These results are relevant for environmental management because they show that the system can be used as a reliable tool to support continuous monitoring and classification of pollution levels in urban contexts. In addition, this level of performance supports its use in applications where reducing false alarms without losing detection capacity is required, such as early warning systems, environmental monitoring dashboards, and decision support platforms. Finally, the results suggest that with improvements in calibration, dataset enrichment, and model adjustment, the system could scale to more complex scenarios and strengthen its usefulness in smart cities and other environments where real time environmental monitoring is strategically important.

Figure 11 shows a confusion matrix that allows the evaluation of the performance of the IoT-based intelligent system in classifying air pollution categories in Seoul. This representation shows both the correct classifications and the model's errors between classes, allowing the analysis of its actual capacity to differentiate atmospheric pollutants with similar environmental behavior.

doi: 10.13053/CyS-30-1-6346

ISSN 2007-9737

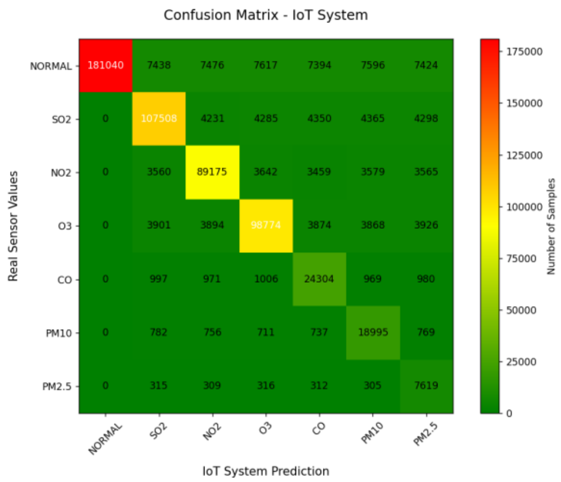


Fig. 11. Confusion matrix of the IoT system

As observed in the results, the main diagonal concentrates the highest values of correct classifications, confirming the model's capability to recognize the different classes. The NORMAL category records 181,040 correct predictions, considerably higher than the other classes, which can be explained by its greater presence in the dataset and the stability of its environmental patterns. The remaining categories also show high values on the diagonal, including 107,508 for SO_2, 89,175 for NO_2, 90,774 for O_3, 24,304 for CO, 18,995 for PM10, and 7,619 for PM2.5, indicating strong overall system performance.

However, some confusion appears between pollutants with similar environmental behaviors. For instance, the NORMAL class is confused with other categories in values close to 7,400 records, suggesting that certain environmental conditions may exhibit intermediate characteristics between normal states and pollution episodes. Likewise, pollutants such as SO_2 and NO_2 present cross-classification errors of approximately 3,500 to 4,300 cases, which may be attributed to common emission sources such as vehicular traffic or urban combustion processes that generate correlations in the concentrations measured by the sensors.

The results obtained from the confusion matrix of the IoT system are consistent with evidence reported in recent studies. First, Alamro et al. [2] demonstrate that machine learning models applied in IoT environments typically achieve high classification accuracy, with most predictions concentrated along the main diagonal, which aligns with the behavior observed in our system. Similarly, Alsamrai et al. [3] report that intelligent environmental monitoring systems tend to present minimal errors between closely related categories due to the natural overlap of measurements, precisely as reflected in the minor confusions observed in our matrix. Finally, Yildiz and Sucuoglu [36] highlight that IoT-based forecasting and classification systems maintain stable performance even under data variability, achieving correct class discrimination with low error rates. These findings support the consistency and reliability of the predictive behavior demonstrated by our model. Collectively, these studies confirm that the performance shown in the confusion matrix of the developed system is consistent with expectations for modern IoT platforms integrating classification algorithms and real-time monitoring.

These results indicate that the IoT system demonstrates adequate performance to support continuous monitoring of urban air pollution. Furthermore, the confusion matrix helps identify pollutants that may require additional calibration or model refinement. These findings support the use of the system in environmental management platforms and early warning systems. Additionally, the proposed approach could be replicated in other cities equipped with IoT sensor networks adapted to their local environmental conditions.

5.2 Spatiotemporal Hotspot Analysis of Urban Air Pollution in Seoul

Figures 12 and 13, together with Table 3, present an integrated spatial analysis of the average PM2.5 concentration across monitoring stations in the city of Seoul. This combined representation enables the identification of persistent exposure differences among districts and facilitates the detection of urban air pollution hotspots.

The results indicate that Yeongdeungpo-gu (29.2), Mapo-gu (28.9), Gwanak-gu (28.0), and Guro-gu (27.6) concentrate the highest average PM2.5 values, exceeding the overall network average (25.0). This suggests localized environmental pressure associated with higher traffic intensity, mixed urban land use, logistics corridors, and possibly a greater density of nearby emission sources.

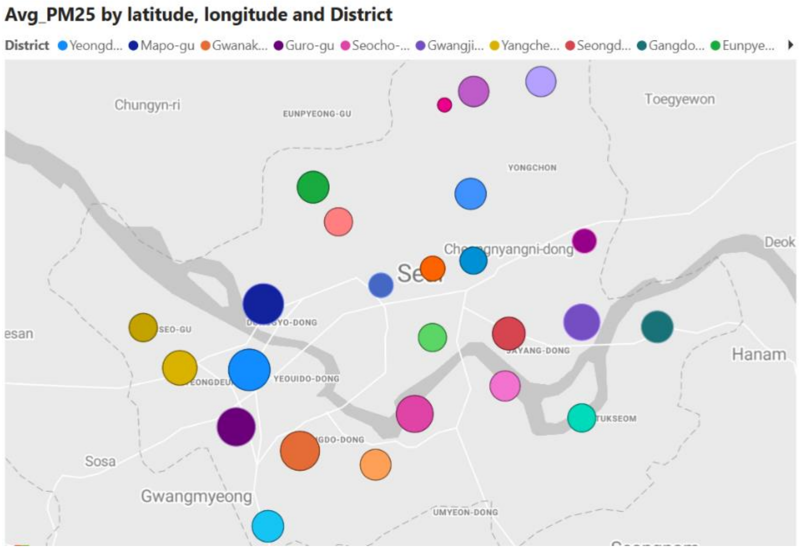


Fig. 12. Spatial distribution of average PM2.5 concentrations across monitoring stations in Seoul. Circle size represents the magnitude of the average PM2.5 concentration recorded at each station

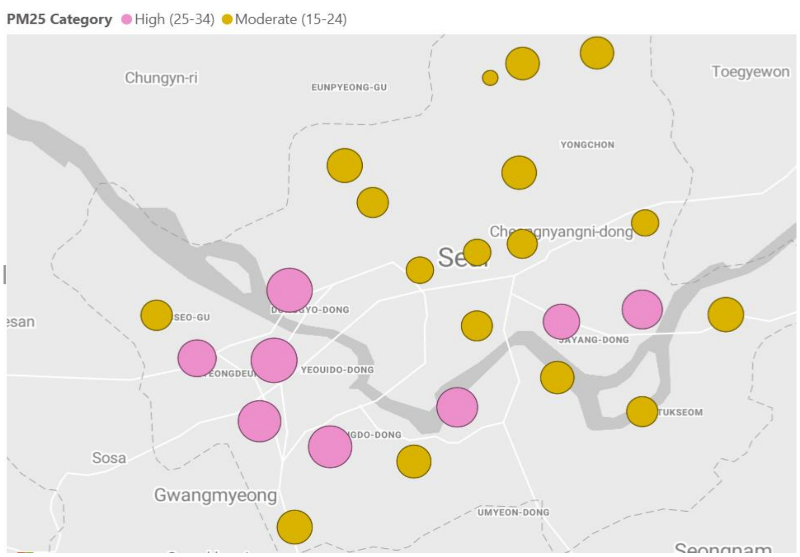


Fig. 13. Spatial distribution of PM2.5 pollution categories across monitoring stations in Seoul. Circle size represents the average PM2.5 concentration, while colors indicate pollution categories

According to the findings presented, spatial heterogeneity is not random, since the highest values are mainly clustered in western and central zones, while Gangbuk-gu records the lowest average (21.8).

This pattern indicates that urban morphology, the uneven distribution of economic activities, and differences in local ventilation may influence the accumulation and dispersion of particulate matter.

Based on the observed distribution, the proximity of several critical points to highly mobile urban corridors and areas of intense metropolitan interaction reinforces the hypothesis that transportation dynamics, traffic congestion, and

ISSN 2007-9737

Table 3. Average PM2.5 concentration by District in Seoul

District	Station Code	Avg PM2.5
Yeongdeungpo-gu	119	29.2
Mapo-gu	106	28.9
Gwanak-gu	121	28.0
Guro-gu	117	27.6
Seocho-gu	122	26.9
Gwangjin-gu	108	26.7
Yangcheon-gu	115	25.9
Seongdong-gu	107	25.2
Gangdong-gu	125	25.0
Eunpyeong-gu	104	24.9
Geumcheon-gu	118	24.8
Seongbuk-gu	111	24.7
Dongjak-gu	120	24.6
Dobong-gu	113	24.4
Gangnam-gu	123	24.4
Nowon-gu	114	24.3
Gangseo-gu	116	23.8
Seodaemun-gu	105	23.8
Songpa-gu	124	23.8
Yongsan-gu	103	23.7
Dongdaemun-gu	109	23.5
Jung-gu	102	22.9
Jongno-gu	101	22.9
Jungnang-gu	110	22.9
Gangbuk-gu	112	21.8
Total (Average)		**25.0**

particle resuspension contribute consistently to elevated average concentrations beyond isolated pollution events.

In line with the identified spatial pattern, the 7.4-point difference between the maximum (29.2) and minimum (21.8) values reveals a relevant intra-urban gradient from a public health perspective, demonstrating that a single city does not experience homogeneous exposure. Consequently, decisions based solely on global averages may underestimate urban microenvironments with comparatively higher risk.

In light of the recorded behavior, the IoT system demonstrates value not only as a continuous monitoring mechanism but also as an analytical infrastructure for detecting persistent hotspots and prioritizing targeted interventions. The stability of the station averages suggests that structural territorial factors—rather than only temporal variability—play a decisive role in shaping the pollution map. The classification of monitoring stations into moderate and high pollution categories confirms that a significant proportion of districts exceed the 25 µg/m³ threshold, evidencing the existence of persistent urban hotspots and revealing that environmental exposure within the city is not homogeneous.

These findings can be extended to other sectors and business areas such as urban logistics, intelligent transportation, insurance, occupational health, construction, and real estate management.

The convergence of IoT and advanced analytics has been successfully applied across multiple domains where continuous monitoring and real-time data analysis are required [42], as noted by Andrade-Mogollón. Likewise, this approach is transferable to other geographic regions—especially large Latin American or Asian metropolitan areas—and to different time periods, enabling comparisons of spatial changes before and after environmental policies, seasonal variations, or urban infrastructure transformations. Finally, the use of IoT networks combined with spatial analytics strengthens evidence-based urban planning, improves the targeting of mitigation measures, and opens the possibility of developing predictive models and early warning systems with multisectoral applications.

Figures 14 and 15, together with Table 4, present the temporal dynamics of PM2.5 concentration at daily and monthly scales, along with future projections derived from a time-series model with a 95% confidence interval.

Complementarily, the table allows examination of the temporal and spatial distribution of pollution by year and district, providing an integrated view of

ISSN 2007-9737

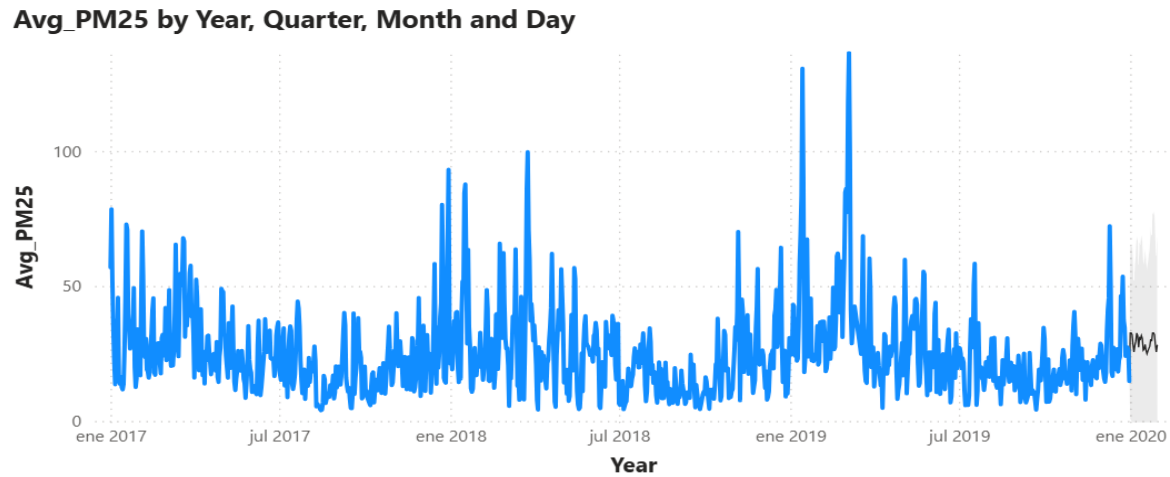


Fig. 14. Daily Average PM2.5 concentrations and 30-Period Forecast with 95% Confidence Interval

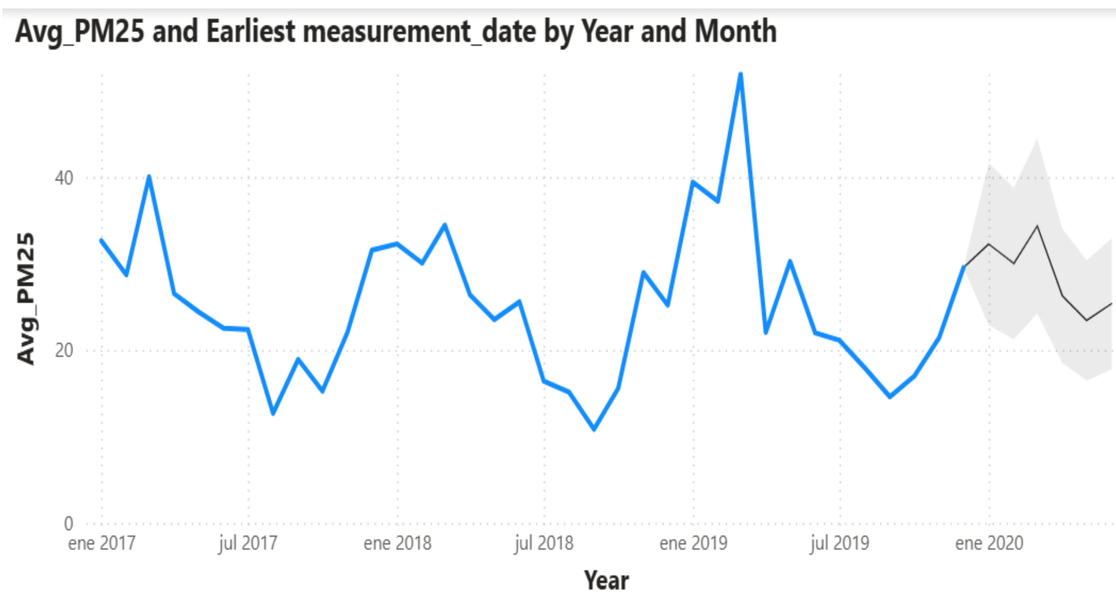


Fig. 15. Monthly Average PM2.5 concentration and Six-Period Forecast with 95% Confidence Interval

the chronological and territorial variation patterns of particulate matter.

As observed in the results, the daily time series shows that between January 2017 and the end of 2019 PM2.5 concentrations exhibit high variability, with pollution episodes exceeding 100 µg/m³, particularly around early 2018 and early 2019. This suggests the occurrence of specific high-emission events or atmospheric conditions that favor the accumulation of particulate matter. According to the findings presented, during the 2017–2018 period moderate values dominate, fluctuating approximately between 10 and 40 µg/m³, whereas by 2019 more pronounced peaks are observed. These variations may be associated with increases in urban activity, vehicular congestion, or episodes

Table 4. Average PM2.5 concentrations by monitoring station and district in Seoul

Year	District	Avg PM2.5
2019	Yeongdeungpo-gu	33.2
2019	Guro-gu	32.2
2019	Gwangjin-gu	31.3
2019	Seocho-gu	31.0
2019	Mapo-gu	30.9
2018	Yeongdeungpo-gu	30.7
2019	Gwanak-gu	29.5
2017	Yangcheon-gu	28.6
2018	Gwanak-gu	28.3
2017	Mapo-gu	28.0
2018	Mapo-gu	27.9
2019	Gangdong-gu	27.1
2019	Seongdong-gu	26.8
2018	Eunpyeong-gu	26.5
2019	Dongjak-gu	26.4
2019	Seongbuk-gu	26.4
2019	Dobong-gu	26.3
2017	Gwanak-gu	26.2
2019	Yangcheon-gu	26.2
2019	Nowon-gu	26.0
2018	Seocho-gu	25.9
2018	Guro-gu	25.8
2017	Geumcheon-gu	25.8
2017	Seongbuk-gu	25.7
2019	Gangnam-gu	25.7
2019	Gangseo-gu	25.5
2017	Dongjak-gu	25.5
2019	Geumcheon-gu	25.5
2017	Nowon-gu	25.4
2017	Gangnam-gu	25.3
Total		**25.0**

of atmospheric stability that reduce pollutant dispersion.

Based on the distribution observed in the monthly series, the average concentrations reveal a cyclical pattern, with notable increases around January and February 2019, where the monthly average exceeds approximately 50 µg/m³, followed by progressive declines toward the middle of the same year. This pattern suggests the influence of seasonal factors, such as changes in temperature, atmospheric circulation, or urban mobility dynamics. The forecasting model indicates that from January 2020 onward, projected concentrations would tend to remain within an approximate range of 20 to 35 µg/m³ in both the short-term horizon (30 days) and the six-month monthly forecast, suggesting a relative stabilization of pollution levels in the absence of extraordinary disturbances.

From a critical perspective, the width of the confidence intervals in the forecasted periods indicates that the dynamics of PM2.5 remain subject to significant uncertainty, since external variables such as meteorological conditions, traffic intensity, or industrial emissions can abruptly alter the trajectory projected by the time-series model.

These results provide useful evidence for the design of environmental monitoring systems, urban mobility management, and public health policies aimed at mitigating atmospheric pollution episodes. Furthermore, the methodology employed can be replicated in other cities or metropolitan regions, as well as applied to different temporal periods, enabling comparisons of pollution patterns across diverse geographic contexts. Finally, this type of analysis is also relevant for business sectors such as transportation, logistics, energy, construction, and insurance, where air quality information can support operational decisions, strategic planning, and environmental risk assessment.

5.3 Experimental Comparison Between GC and GE

5.3.1 Experimental Results

Finally, measurements were carried out and 30 observations were collected for additional indicators, including the reduction in update time from the moment information is entered into the database until its visualization in the reports, the

Table 5. Post-test results of indicators for the control group (Gc) and experimental group (Ge)

	Report generation time (s)		% of Incorrect data		% of Missing data	
N°	**Gc**	**Ge**	**Gc**	**Ge**	**Gc**	**Ge**
1	552.34	6.30	0.742	0.063	2.42	1.37
2	558.12	6.37	1.407	0.066	3.28	1.42
3	549.89	6.59	0.514	0.0628	2.61	1.38
4	533.45	6.41	1.484	0.0633	3.09	1.52
5	569.22	6.55	0.826	0.0607	2.71	1.44
6	542.77	6.68	0.860	0.0556	2.15	1.40
7	564.03	6.25	1.062	0.0541	2.83	1.40
8	557.68	6.53	1.291	0.0607	2.41	1.35
9	536.54	6.60	1.029	0.0551	2.63	1.43
10	528.91	6.45	0.517	0.0602	2.44	1.46
11	551.40	6.49	1.187	0.0669	3.01	1.43
12	545.66	6.66	1.187	0.0468	2.52	1.41
13	568.29	6.35	0.650	0.0515	2.50	1.48
14	525.84	6.39	1.352	0.0535	2.47	1.46
15	560.17	5.92	0.917	0.0566	2.58	1.41
16	539.72	6.27	1.142	0.0582	2.30	1.52
17	547.09	6.20	0.639	0.0613	2.28	1.39
18	571.31	6.46	1.445	0.0479	2.60	1.28
19	554.48	6.22	1.083	0.0571	2.55	1.42
20	530.62	6.31	1.365	0.0654	2.41	1.29
21	543.15	6.48	1.011	0.0618	2.19	1.30
22	548.93	6.40	0.775	0.0592	2.04	1.45
23	562.07	6.54	0.852	0.0566	2.62	1.47
24	555.52	6.19	0.935	0.0592	3.11	1.40
25	541.28	6.34	1.311	0.0577	2.66	1.37
26	526.47	6.15	1.247	0.0582	2.32	1.44
27	553.36	6.28	0.991	0.0705	2.33	1.42
28	537.58	6.23	1.052	0.0561	2.38	1.51
29	550.11	6.42	1.445	0.0592	2.12	1.30
30	544.98	6.21	1.487	0.0664	2.20	1.47

decrease in the percentage of incorrect data in the records, and the reduction of missing data associated with null values, as shown in Table 5. Their inclusion is methodologically relevant because it allows, through repeated observations, the evaluation of the actual effect of implementing the IoT-based intelligent system on processing efficiency and data quality.

5.3.2 Normality Test

The normality test applied to the experimental data showed that the observations follow an approximately normal distribution, as illustrated in Figures 16, 17, and 18, where the data present an adequate alignment with the expected normality pattern. This result allows the assumption of normality in the analyzed indicators and justifies the application of parametric statistical methods to compare the results between the experimental group (Ge) and the control group (Gc).

– **I1**. Report generation time

It is observed that, for both the control group (Gc) and the experimental group (Ge) in the post-test for

ISSN 2007-9737

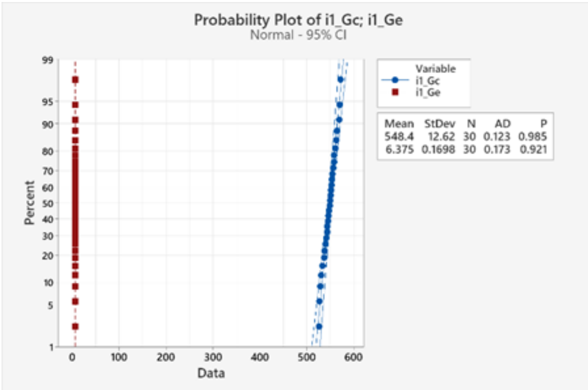


Fig. 16. Normality test for Indicator 1

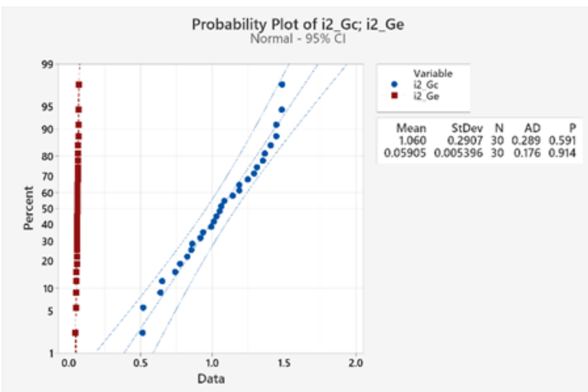


Fig. 17. Normality test for Indicator 2

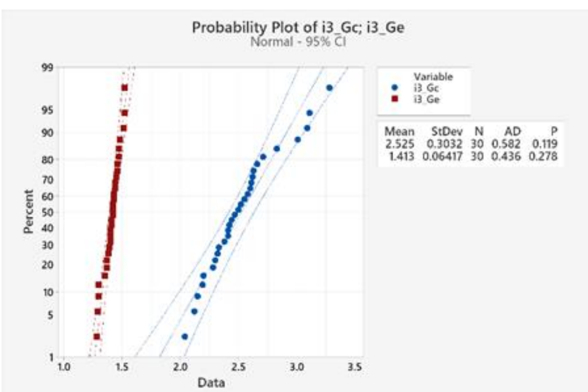


Fig. 18. Normality test for Indicator 3

this indicator, the obtained p-values (0.985 and 0.921, respectively) are greater than the significance level α = 0.05. Consequently, the data for this indicator exhibit behavior consistent with a normal distribution.

– **I2.** % of Incorrect data

It can be observed that, for both the control group (Gc) and the experimental group (Ge) in the post-test for this indicator, the obtained p-values (0.591 and 0.914, respectively) are greater than the significance level α = 0.05. Consequently, the data for this indicator exhibit behavior consistent with a normal distribution.

– **I3.** % of Missing data

It is observed that, for both the control group (Gc) and the experimental group (Ge) in the post-test for this indicator, the obtained p-values (0.119 and 0.278, respectively) are greater than the significance level α = 0.05. Therefore, the data for this indicator exhibit behavior consistent with a normal distribution.

5.3.3 Descriptive Statistics and Distribution Analysis

To complement the normality assessment, a descriptive statistical analysis was performed using Minitab software in order to characterize the distribution of the indicators obtained in the post-test. This analysis includes measures of central tendency, dispersion, confidence intervals, kurtosis, and skewness, enabling a more precise evaluation of the stability and consistency of the results obtained for each indicator, as shown in Tables 6 and 7.

The results show that the Anderson–Darling normality test produced AD and p-values greater than α (0.05); therefore, the normality of the data for the analysis was confirmed. It was also observed that, with a 95% confidence level, the mean and standard deviation revealed normal results in the data corresponding to the research indicators.

The results show that the three indicators present narrow 95% confidence intervals, indicating that the means are estimated with good precision. The kurtosis values are close to zero, which indicates the absence of abnormal peaks or heavy tails. The skewness values are slightly negative, suggesting a mild tendency toward values lower than the mean. The third quartile (Q3) indicates that 75% of the values are less than or

Table 6. Descriptive statistics and Anderson–Darling normality results for post-test indicators

Sample	n	Mean	StDev	AD	p-value
I1: Post-test (Gc)	30	548.4	12.62	0.123	0.985
I1: Post-test (Ge)		6.375	0.1698	0.173	0.921
I2: Post-test (Gc)	30	1.060	0.2907	0.289	0.591
I2: Post-test (Ge)		0.05905	0.005396	0.176	0.914
I3: Post-test (Gc)	30	2.525	0.3032	0.582	0.119
I3: Post-test (Ge)		1.413	0.06417	0.436	0.278

Table 7. Confidence intervals and distribution shape statistics for experimental-group indicators

Sample	n	95% confidence interval for the mean	Kurtosis	Skewness
I1: Post-test (Ge)	30	6.3113 – 6.4381 s	0.35	-0.32
I2: Post-test (Ge)	30	0.057038 – 0.061068 %	0.27	-0.22
I3: Post-test (Ge)	30	1.3890 – 1.4370 %	-0.09	-0.45

equal to this value, demonstrating a consistent distribution without significant outliers.

– **Indicator I1: Report generation time**

The comparison between Gc and Ge shows a drastic reduction in the average report generation time, from 548.4 s to 6.375 s, with standard deviations of 12.62 and 0.1698, respectively. This result demonstrates not only greater speed but also significantly lower operational variability in the experimental system. The AD values (0.123 in Gc; 0.173 in Ge) and p-values (0.985 and 0.921) confirm normality in both groups, while the 95% CI of Ge (6.3113–6.4381 s) indicates a highly precise estimation of the mean. The kurtosis (0.35) and skewness (−0.32) suggest a stable distribution slightly biased toward lower times, which can be attributed to the automation of the data flow, the elimination of manual tasks, and the direct synchronization between data capture, processing, and visualization.

The results obtained for indicator I1 are consistent with recent literature that attributes substantial reductions in environmental data processing times to IoT technologies. In this context, Bobulski and colleagues [5] show that IoT systems enable real-time measurements with a significant reduction in operational latency. Similarly, Islam and collaborators [14] demonstrate that integrating IoT with machine learning techniques optimizes the management of large volumes of data, while Zafra-Pérez and their research team [37] highlight that low-cost mobile sensors improve the speed, stability, and continuity of environmental reporting even in complex environments. Likewise, Jo and co-authors [17] and Wang and colleagues [35] emphasize that IoT-based sensor systems can generate reports almost instantaneously even in scenarios with high particulate matter concentrations, which aligns with the significant temporal reduction observed in the experimental group.

In a similar vein, Daffa Pebrian and collaborators [7] and Collado and colleagues [6] report that open IoT platforms reduce latency in continuous monitoring. Moreover, Abdelmalek and co-authors [1], Alsamrai and their research team [3], and Ramadan and collaborators [28] show that the integration of IoT and artificial intelligence not only accelerates data processing but also enhances system scalability without compromising efficiency, while Zhao and colleagues [40] emphasize that these technologies improve operational responsiveness by optimizing report generation. Although some studies are conducted in specific contexts or under different architectures, they converge on a common conclusion: automation of the data flow substantially reduces response times. Consequently, the behavior observed in the present study is methodologically consistent with this evidence.

Conversely, non-automated approaches are typically more exposed to delays associated with manual tasks, fragmented validation processes, and lower operational continuity. Therefore, the drastic reduction in report generation time observed in the experimental group can be interpreted as a direct effect of integrating sensors, connectivity, and automated processing. Overall, the convergence of the evidence supports the validity of the results obtained and confirms that IoT constitutes an effective strategy for improving the timeliness and efficiency of environmental monitoring. IoT systems rely on networks of connected sensors that collect environmental data in real time and transmit them to centralized systems to support timely and efficient decision-making [42], as noted by Andrade-Mogollón and colleagues.

These results imply that the IoT system significantly improves operational efficiency, reducing response time from several minutes to only a few seconds while maintaining highly stable execution. The combination of a low mean, reduced standard deviation, narrow confidence interval, and regular distribution shape supports its application in contexts where reporting timeliness is critical. In practical terms, this strengthens the feasibility of the system for real-time environmental surveillance, alert activation, and rapid decision-making in dynamic urban environments.

– **Indicator I2: Percentage of incorrect data**

Indicator I2 shows a substantial improvement in data quality, as the mean decreases from 1.060% in Gc to 0.05905% in Ge, accompanied by a significant reduction in dispersion (StDev = 0.2907 vs. 0.005396), suggesting greater stability in error control. The normality statistics (AD = 0.289 and 0.176; p-value = 0.591 and 0.914) confirm the parametric behavior of the data, while the 95% CI of Ge (0.057038–0.061068%) indicates a highly concentrated mean. The kurtosis (0.27) close to zero and skewness (−0.22) indicate a regular distribution slightly inclined toward lower values, likely due to automated validation, cleaning routines, and consistency checks that reduce typing, duplication, or transfer errors.

The results of indicator I2 are consistent with recent literature indicating that IoT systems significantly reduce anomalous records through automated validation, calibration, and data-cleaning mechanisms. Abdelmalek and colleagues [1] show that machine learning models applied to environmental monitoring detect atypical patterns and correct inconsistencies at early stages. Likewise, Banciu and collaborators [4] and Bobulski and their research team [5] highlight that IoT platforms integrate mechanisms for detecting defective values that reduce measurement errors. Similarly, Daffa Pebrian and co-authors [7] and Felici-Castell and colleagues [8] explain that the integration of sensors, stable networks, and distributed processing reduces the probability of data corruption during transmission, while Giordano and collaborators [12] and Islam and their research team [14] show that internal validation algorithms improve the coherence of IoT-generated records.

Furthermore, Jaramillo-Perez and colleagues [16] indicate that intelligent filtering and automated analysis allow anomalous records to be removed without manual intervention. Karnati [19] and Khan and collaborators [20] demonstrate that continuous sensor calibration reduces errors in real monitoring contexts. In addition, Mahajan and Helbing [23] and Sarmiento Sánchez [30] report that redundant validation and noise control significantly improve data quality in IoT networks. Temkov and co-authors [32] and Veiga and colleagues [34] also highlight that properly configured networks and robust transmission protocols reduce failures associated with incorrect records. Finally, Wang and collaborators [35] and Zaid and their research team [39] demonstrate that cross-calibration, intelligent analysis, and continuous supervision contribute to minimizing errors in dynamic air-quality monitoring. Although the application contexts vary, all studies converge in demonstrating that automated data validation significantly reduces operational errors. Consequently, the reduction observed in the experimental group can be interpreted as a direct effect of integrating IoT sensors, analytical algorithms, and data-quality control mechanisms.

The reduction of error rates and their low variability indicate a clear improvement in dataset integrity, increasing the reliability of subsequent analyses and predictive models built on these records. The confirmation of normality, the narrow confidence interval, and the stable distribution shape suggest that the reduction in errors is not episodic but consistent over time. This makes the system particularly useful in applications where data accuracy is essential, such as continuous monitoring, environmental health analysis, energy management, and industrial sensor networks.

– **Indicator I3: Percentage of missing data**

For I3, the average percentage of missing data decreases from 2.525% in Gc to 1.413% in Ge, accompanied by a significant reduction in dispersion (0.3032 to 0.06417), reflecting greater completeness and uniformity in the records generated by the experimental system. The AD values (0.582 and 0.436) and p-values (0.119 and 0.278) indicate normality in both groups, and the 95% CI of Ge (1.3890–1.4370%) shows that the mean is estimated with high precision. The kurtosis (−0.09) and skewness (−0.45) reveal a stable distribution without extreme tails and with a slight inclination toward lower values, which may be explained by a more robust architecture for data capture, transmission, and storage, with less dependence on manual processes prone to omissions.

The results for indicator I3 are consistent with recent evidence showing that IoT architectures improve data completeness through continuous validation mechanisms, stable transmission, and redundancy in data acquisition. Islam and colleagues [14] demonstrate that integrating IoT with machine learning models enables real-time detection and correction of null values, reducing information loss. Similarly, Jaramillo-Perez and collaborators [16] highlight that structured environmental data acquisition processes improve record continuity and minimize interruptions during measurement. Karnati [19] shows that IoT platforms with automated analysis incorporate early failure detection mechanisms that reduce incomplete records, while Mahajan and Helbing [23] demonstrate that dynamic sensor calibration reduces data loss during transmission.

Likewise, Felici-Castell and their research team [8] report that intelligent sensor networks operating under robust architectures maintain very low levels of missing records even under high environmental variability. Zafra-Pérez and co-authors [38] also explain that optimized wireless sensor networks allow continuous and stable real-time data capture. Finally, Zaid and collaborators [39] show that low-cost IoT systems with integrated intelligent analytics can efficiently manage redundancy and data recovery, preventing significant information loss. Although methodological approaches differ across studies, they consistently conclude that automation of data acquisition and validation substantially reduces missing records. Consequently, the decrease observed in the experimental group can be interpreted as a direct effect of integrating sensors, verification algorithms, and robust transmission mechanisms.

The lower mean of missing data, together with its low dispersion and stable distribution, indicates that the experimental system consistently improves the completeness of environmental records. This is crucial because reducing data gaps strengthens analytical continuity, reduces the need for data imputation, and improves the quality of time-series analyses and predictive models. Consequently, the system demonstrates potential for scaling to scenarios where continuous information availability is critical, including urban monitoring, smart infrastructure management, and decision-support platforms based on continuous data streams. As future work, it is proposed to strengthen the system architecture by prioritizing the identification and mitigation of vulnerabilities in IoT objects, focusing on privacy, access control, and data storage, adopting a comprehensive cybersecurity strategy as recommended in recent studies.

6 Conclusions and Future Research

The findings of the study confirm that the IoT-based intelligent system substantially improved operational performance and data quality compared with the control group. First, I1 (report generation time) showed the most pronounced difference, reducing the mean from 548.4 s in the control group to 6.375 s in the experimental group, with markedly lower dispersion, demonstrating a simultaneous improvement in both speed and process stability. Likewise, I2 (% of incorrect data) decreased from 1.060% to 0.05905%, indicating that the automation of the validation and processing workflow contributed to minimizing errors and strengthening record integrity. In addition, I3 (% of missing data) declined from 2.525% to 1.413%, reflecting a sustained improvement in data completeness and reduced dependence on manual processes prone to omissions. Across all three indicators, the normality results, the narrow confidence intervals, and the dispersion statistics support the consistency of the observed behavior.

Overall, these results lead to the conclusion that the proposed architecture not only optimizes report generation but also enhances the reliability and continuity of environmental information, which are essential aspects for real-time urban monitoring. Therefore, the developed system represents a technically robust alternative for strengthening environmental surveillance processes, supporting data-driven decision-making, and scaling toward smart city contexts where the timeliness, quality, and availability of data are critical conditions for air pollution management.

As a future research direction, it is recommended to expand the system architecture through advanced cybersecurity mechanisms, access control policies, and secure data storage protection, as well as incorporating additional environmental variables, adaptive calibration strategies, and validation in other cities or urban contexts. These improvements would strengthen the robustness, scalability, and interoperability of the system in more complex operational environments.

References

1. **Abdelmalek, M.M., Mahmoud, H., Shokry, H. (2025).** Prognosis of Air Quality Index and Air Pollution Using Machine Learning Techniques. Scientific Reports, Vol. 15, No. 1. doi: 10.1038/s41598-025-11260-y.
2. **Alamro, H., Al-Wesabi, F.N., Alahmari, S., Aljabri, J., Ebad, S.A., Alshuhail, A., Alallah, F.S., Alshareef, A., Sharif, M.M. (2025).** Feature Enhancement Model with Up Sampling Based Cyber Threat Attack Detection and Classification on Imbalanced Dataset in Industrial Internet of Things. Alexandria Engineering Journal, Vol. 128, pp. 247–258. doi: 10.1016/j.aej.2025.05.060.
3. **Alsamrai, O., Redel-Macias, M.D., Dorado, M.P. (2025).** Real-Time Intelligent Monitoring of Outdoor Air Quality in an Urban Environment Using IoT and Machine Learning Algorithms. Applied Sciences, Vol. 15, No. 16. doi: 10.3390/app15169088.
4. **Banciu, C., Florea, A., Bogdan, R. (2024).** Monitoring and Predicting Air Quality with IoT Devices. Processes, Vol. 12, No. 9. doi: 10.3390/pr12091961.
5. **Bobulski, J., Szymoniak, S., Pasternak, K. (2024).** An IoT System for Air Pollution Monitoring with Safe Data Transmission. Sensors, Vol. 24, No. 2. doi: 10.3390/s24020445.
6. **Collado, E., Calderón, S., Cedeño, B., de León, O., Centella, M., García, A., Sáez, Y. (2024).** Open-Source Internet of Things (IoT)-Based Air Pollution Monitoring System with Protective Case for Tropical Environments. HardwareX, Vol. 19. doi: 10.1016/j.ohx.2024.e00560.
7. **Daffa Pebrian, D., Delphiano, A.H., Diniarti, S., Aly Afandi, M., Ginting, M.B. (2025).** IoT-Based Air Quality Monitoring System Utilizing OpenRemote as an Application Server at University. Buletin Pos Dan Telekomunikasi, Vol. 23, No. 1, pp. 57–75. doi: 10.17933/bpostel.v23i1.412.
8. **Felici-Castell, S., Segura-Garcia, J., Perez-Solano, J.J., Fayos-Jordan, R., Soriano-Asensi, A., Alcaraz-Calero, J.M. (2023).** AI-IoT Low-Cost Pollution-Monitoring Sensor Network to Assist Citizens with Respiratory Problems. Sensors, Vol. 23, No. 23. doi: 10.3390/s23239585.
9. **Fikri, A., Ridwan, M., Agustine, D. (2025).** Implementation of Internet of Things (IoT) in Air Quality Monitoring (AQI). Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer, Vol. 5, No. 2, pp. 424–443. doi: 10.55606/juisik.v5i2.1326.
10. **Fu, Y., Lu, H., Chen, J., Luo, B. (2025).** Key–Value Data Collection with Local Differential Privacy for Urban Air Quality Monitoring in Crowdsensing. Internet of Things, Vol. 33. doi: 10.1016/j.iot.2025.101670.
11. **Gamboa-Cruzado, J., Blanco-Maslucan, R., Galindo-Arone, D., Jáuregui-Romero, H., Meza, A.N., Arauco-Esquivel, S., Núñez-**

Prado, C.J. (2025). Augmented Reality as a Tool for Cultural Learning in Educational Institutions in Peru. Computación y Sistemas, Vol. 29, No. 2, pp. 901–915. doi: 10.13053/CyS-29-2-5739.

12. **Giordano, T., Brilli, L., Gualtieri, G., Martelli, F., Cavaliere, A., Carotenuto, F., Nardino, M., Fiorillo, E., Zaldei, A., Putzolu, S., Vagnoli, C., Castelli, F., Gioli, B. (2025).** Assessment of Risk Components for Urban Population to Heat Intensity and Air Pollution through a Dense IoT Sensor Network. Urban Climate, Vol. 61. doi: 10.1016/j.uclim.2025.102397.
13. **Hassan, A.K., Saraya, M.S., Ali-Eldin, A.M.T., Abdelsalam, M.M. (2024).** Low-Cost IoT Air Quality Monitoring Station Using Cloud Platform and Blockchain Technology. Applied Sciences, Vol. 14, No. 13. doi: 10.3390/app14135774.
14. **Islam, M.M., Jibon, F.A., Tarek, M.M., Kanchan, M.H., Perbhez Shakil, S.U. (2024).** A Real-Time Dataset of Air Quality Index Monitoring Using IoT and Machine Learning in the Perspective of Bangladesh. Data in Brief, Vol. 55. doi: 10.1016/j.dib.2024.110578.
15. **Jang, B.J., Park, N., Jung, I. (2025).** IoT Sensing-Based High-Density Monitoring of Urban Roadside Particulate Matter (PM10 and PM2.5). Applied Sciences, Vol. 15, No. 21. doi: 10.3390/app152111608.
16. **Jaramillo-Perez, J.M., Macías-Hernández, B.A., Tello-Leal, E. (2025).** A Dataset of Volatile Organic Compounds (VOCs) Concentrations Collected Near a Petroleum Refinery Area. Data in Brief, Vol. 61. doi: 10.1016/j.dib.2025.111900.
17. **Jo, J.H., Jo, B.W., Kim, J.H., Choi, I. (2020).** Implementation of IoT-Based Air Quality Monitoring System for Investigating Particulate Matter (PM10) in Subway Tunnels. International Journal of Environmental Research and Public Health, Vol. 17, No. 15, pp. 1–12. doi: 10.3390/ijerph17155429.
18. **Johnson, T. (2025).** Digital Exposome: A Dataset for Wellbeing Classification Using Environmental Air Quality and Human Physiological Data. Data in Brief, Vol. 59. doi: 10.1016/j.dib.2025.111442.
19. **Karnati, H. (2023).** IoT-Based Air Quality Monitoring System with Machine Learning for Accurate and Real-Time Data Analysis. arXiv e-prints, arXiv:2307.00580. doi: 10.48550/arXiv.2307.00580.
20. **Khan, M.A., Kim, H.C., Park, H. (2022).** Leveraging Machine Learning for Fault-Tolerant Air Pollutants Monitoring for a Smart City Design. Electronics, Vol. 11, No. 19. doi: 10.3390/electronics11193122.
21. **Kim, Y.-H., Moon, S.-H. (2025).** Machine Learning-Based Quality Control for Low-Cost Air Quality Monitoring: A Comprehensive Review of the Past Decade. Atmosphere, Vol. 16, No. 10, pp. 1136. doi: 10.3390/ atmos16101136.
22. **Lolli, S. (2025).** Urban PM2.5 Concentration Monitoring: A Review of Recent Advances in Ground-Based, Satellite, Model, and Machine Learning Integration. Urban Climate, Vol. 63. doi: 10.1016/j.uclim.2025.102566.
23. **Mahajan, S., Helbing, D. (2025).** Dynamic Calibration of Low-Cost PM2.5 Sensors Using Trust-Based Consensus Mechanisms. Npj Climate and Atmospheric Science, Vol. 8, No. 1. doi: 10.1038/s41612-025-01145-2.
24. **Parkavi, A., Sowmya, B.J., Alex, S.A., Supreeth, S., Shruthi, G., Rohith, S., Chatterjee, S., Lingaraj, K. (2025).** Air Quality and Dust Level Monitoring Systems in Hospitals Using IoT. Discover Internet of Things, Vol. 5, No. 1. doi: 10.1007/s43926-025-00120-w.
25. **Parra-Medina, F.D., Vélez-Guerrero, M.A., Callejas-Cuervo, M. (2024).** Low-Cost Solution for Air Quality Monitoring: Unmanned Aerial System and Data Transmission via LoRa Protocol. Sustainability, Vol. 16, No. 22. doi: 10.3390/su162210108.
26. **Pati, A.K., Tripathy, A.R., Subudhi, S. (2026).** Intelligence Frameworks for Environmental Pollution Assessment: A Review on Air and Water Quality Monitoring Systems. Measurement, Vol. 258. doi: 10.1016/j.measurement.2025.119122.
27. **Rakholia, R., Le, Q., Vu, K., Ho, B.Q., Carbajo, R.S. (2022).** Outdoor Air Quality Data for Spatiotemporal Analysis and Air Quality

Modelling in Ho Chi Minh City, Vietnam. Urban Climate, Vol. 46. doi: 10.1016/j.uclim. 2022.101315.

28. **Ramadan, M.N.A., Ali, M.A.H., Khoo, S.Y., Alkhedher, M., Alherbawi, M. (2024).** Real-Time IoT-Powered AI System for Monitoring and Forecasting of Air Pollution in Industrial Environment. Ecotoxicology and Environmental Safety, Vol. 283. doi: 10.1016/j.ecoenv.2024.116856.
29. **Ramos-Romero, A., Garcia-Yataco, B., Andrade-Arenas, L. (n.d.).** Mobile Application Design with IoT for Environmental Pollution Awareness. IJACSA International Journal of Advanced Computer Science and Applications, Vol. 12, No. 1.
30. **Sarmiento Sánchez, C.A. (2025).** Diseño de un nodo sensor Edge IoT para el monitoreo de contaminantes atmosféricos. Revista de Investigación en Tecnologías de la Información, Vol. 13, No. 30, pp. 25–40. doi: 10.36825/riti.13.30.003.
31. **Selvakumar, M., Geetha, S. (2025).** Air Quality Assessment and Validation – Integrating Sensor Network with Line Source Modeling. Global Nest Journal, Vol. 27, No. 5. doi: 10.30955/gnj.06372.
32. **Temkov, S., Cavkovski, P., Lameski, P., Zdravevski, E., Herzog, M.A., Trajkovik, V. (2025).** Air Pollution Data: A Dataset Gathered Through a Crowd Sensing Platform. Data in Brief, Vol. 61. doi: 10.1016/j.dib.2025.111683.
33. **van Poppel, M., Schneider, P., Peters, J., Yatkin, S., Gerboles, M., Matheeussen, C., Bartonova, A., Davila, S., Signorini, M., Vogt, M., Dauge, F.R., Skaar, J.S., Haugen, R. (2023).** SensEURCity: A Multi-City Air Quality Dataset Collected for 2020/2021 Using Open Low-Cost Sensor Systems. Scientific Data, Vol. 10, No. 1. doi: 10.1038/s41597-023-02135-w.
34. **Veiga, T., Munch-Ellingsen, A., Papastergiopoulos, C., Tzovaras, D., Kalamaras, I., Bach, K., Votis, K., Akselsen, S. (2021).** From a Low-Cost Air Quality Sensor Network to Decision Support Services: Steps Towards Data Calibration and Service Development. Sensors, Vol. 21, No. 9. doi: 10.3390/s21093190.
35. **Wang, Z., Chen, Z., Shahid, I., Asif, Z., Haghighat, F. (2025).** Indoor Air Quality Assessment Through IoT Sensor Technology: A Montreal–Qatar Case Study. Atmosphere, Vol. 16, No. 5. doi: 10.3390/atmos16050574.
36. **Yildiz, O., Sucuoglu, H.S. (2025).** Development of Real-Time IoT-Based Air Quality Forecasting System Using Machine Learning Approach. Sustainability, Vol. 17, No. 19, pp. 8531. doi: 10.3390/su17198531.
37. **Zafra-Pérez, A., Boente, C., de la Campa, A.S., Gómez-Galán, J.A., de la Rosa, J.D. (2023).** A Novel Application of Mobile Low-Cost Sensors for Atmospheric Particulate Matter Monitoring in Open-Pit Mines. Environmental Technology and Innovation, Vol. 29. doi: 10.1016/j.eti.2022.102974.
38. **Zafra-Pérez, A., Medina-García, J., Boente, C., Gómez-Galán, J.A., Sánchez de la Campa, A., de la Rosa, J.D. (2024).** Designing a Low-Cost Wireless Sensor Network for Particulate Matter Monitoring: Implementation, Calibration, and Field-Test. Atmospheric Pollution Research, Vol. 15, No. 9. doi: 10.1016/j.apr.2024.102208.
39. **Zaid, M., Nawale, P., Kumar, V., Malyan, V., Sahu, M. (2025).** Investigating IoT-Based Low-Cost Sensor Network for Real-Time Hyper-Local Air Quality Monitoring and Exposure Assessment. Atmospheric Pollution Research. doi: 10.1016/j.apr.2025.102749.
40. **Zhao, Y.L., Tang, J., Huang, H.P., Wang, Z., Chen, T.L., Chiang, C.W., Chiang, P.C. (2020).** Development of IoT Technologies for Air Pollution Prevention and Improvement. Aerosol and Air Quality Research, Vol. 20, No. 12, pp. 2874–2888. doi: 10.4209/aaqr. 2020.05.0255.
41. **Padilla-Gomez, M., Gamboa-Cruzado, J., Távora-Aponte, S., Núñez-Meza, A., Anayo-Gamboa, F., Salazar-Osorio, E. (2024).** Systematic Literature Review on Cybersecurity and Its Influence on Cyber Attacks Targeting IoT Devices. Computación y Sistemas, Vol. 28, No. 4, pp. 1847–1864. doi: 10.13053/CyS-28-4-5006.
42. **Andrade-Mogollon, T., Gamboa-Cruzado, J., Amayo-Gamboa, F. (2025).** Systematic Literature Review of Generative AI and IoT as

Key Technologies for Precision Agriculture. Computación y Sistemas, Vol. 29, No. 2, pp. 857–882. doi: 10.13053/CyS-29-2-5738.

43. **Tupac-Agüero, K., Ortega-Moran, K., Gamboa-Cruzado, J., Menéndez Mueras, R., Del-Valle-Jurado, C., Salazar-Marz, A., Núñez-Meza, A. (2026).** Computer Vision and XRF-IoT Sensor Systems for Detecting Heavy Metals in Export Crops: A Comprehensive Systematic Review. Electronics, Vol. 15, pp. 5062. doi: 10.3390/electronics1505062.

Articles received on 07/11/2025; accepted on 09/01/2026.
**Corresponding author is Obdulia Pichardo-Lagunas.*