



Enhancing Air Quality Monitoring in Mexico City: A Hybrid Sensor-Machine Learning System

Camilo Israel Chávez Galván, Roberto Zagal, Miguel Felix Mata, Fabio Duarte, Simone Mora, Amadeo Arguelles^(✉), and Martina Mazzarello

Av Juan De Dios Batiz S/N, Col Nueva Industrial Vallejo, 07738 Mexico City, Mexico
{alumno2022, rzagal, mmatar, aarguelles}@ipn.mx, {fduarte, moras, mmazz}@mit.edu

Abstract. We present an approach for monitoring and built a dataset of regional historical air quality data in Mexico City. We design a hybrid air quality network prototype that combines mobile and stationary sensors to collect street-level data on particulate matter (PM_{2.5} and PM₁₀). The network is composed of mobile monitoring modules, both stationary at street level and mounted on vehicles, to capture a comprehensive sample of particulate matter behavior in specific areas. Collected data is transmitted using IoT network and processed using machine learning techniques, to generate predictive models to forecast air quality at street level. This approach is an additional improvement to current monitoring capabilities in Mexico City by providing granular street-level data. The system provides a regional and periodic perspective on air quality, enhancing the understanding of pollution levels and supporting informed decision-making to enhance public health and well-being. This research represents a solution for environmental monitoring in urban environments to know how the behavior from pollution levels in air is. The experiments show the effectiveness, and the model of forecast has an overall performance around 81% that is acceptable for the small geographical area testing. As future work is required to include a major number of nodes to collect data from a big geographical coverage and test with other models and algorithms.

Keywords: Air quality · monitoring · IoT · Machine Learning

1 Introduction

Despite decades of global efforts to combat air pollution, the progress seems to be declining due to various factors such as climate change, wildfires, and increased human consumption driven by population growth as other research indicated such as [1]. Mexico City has long been known for its high levels of air pollution [2], and while significant efforts have been made to reduce pollutant emissions, the problem persists not only in the city but also worldwide [4].

One of the most significant challenges in recent decades has been the efforts to reduce the environmental impact we create as a species by polluting our air. It is a phenomenon

that claims thousands of lives annually and increasingly affects large cities. It is no coincidence that there are more and more independent organizations and projects joining the cause to find alternatives to reduce pollution levels on our planet.

For other and, in [3] is used of low-cost air quality monitors (LCMs) and their performance in assessing particulate matter levels. It highlights the advancements in sensor technology that have made LCMs accessible for home use. The study's methodology, which involved comparing LCM measurements to reference data from various sources, is commendable.

Air pollution consists of particles and gases that can reach hazardous concentrations in both outdoor and indoor environments. The adverse effects of high pollution levels range from health risks to increased temperatures. Key pollutants that require constant monitoring include sulfur dioxide (SO₂), carbon monoxide (CO), nitrogen dioxide (NO₂), ozone (O₃), suspended particles (PM₁₀, PM_{2.5}), among others. According to the World Health Organization (WHO), air pollution is estimated to cause 7 million deaths globally each year, with 9 out of 10 people breathing air that exceeds WHO guidelines, disproportionately affecting poorer and developing countries [7].

According to SEDEMA [8], air quality regulations for suspended particles are categorized according to Table 1.

Table 1. Values of air quality in Mexico City.

Contaminant	Values	Quality
PM _{2.5} /PM ₁₀	0–50	Good
PM _{2.5} /PM ₁₀	51–100	Moderate
PM _{2.5} /PM ₁₀	101–150	Poor
PM _{2.5} /PM ₁₀	151–200	Very Poor
PM _{2.5} /PM ₁₀	201–300	Extremely Poor

Similar to other densely populated cities, especially those located in valleys like Mexico City, there are challenges in controlling pollutant levels due to geographical characteristics, particularly with ozone and suspended particles. Although government efforts have gradually reduced emission levels since the 1990s, population growth has hindered further progress in recent years [7]. Currently, Mexico City has the Atmospheric Monitoring System (SIMAT) implemented to comply with regulations for controlling pollutant levels and promptly preventing risks associated with high pollutant concentrations that may endanger public health. SIMAT includes the Automatic Atmospheric Monitoring Network (RAMA), which measures sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone, PM₁₀, and PM_{2.5}. It consists of 34 stations and the Air Quality Information Center (CICA), which serves as the repository for all data generated by the Atmospheric Monitoring System and is responsible for data validation, processing, and dissemination [8]. This network enables the creation of a pollution level map through stationary stations at specific locations. However, despite being in operation for decades,

this monitoring system has not significantly evolved alongside the rapid growth and constant changes of the city.

Our research is motivated by the increasing demand for new technologies like IoT and Machine Learning in problem-solving [5], and how their implementation can greatly improve the lives of people. Inspired by the initiatives of non-governmental organizations in creating intelligent air quality monitoring systems, such as the Environmental Defense Fund in London and research like [6].

This research aims to address the limitations of the current air quality monitoring system in Mexico City by proposing the development of a hybrid network that integrates mobile and stationary sensors with IoT and Machine Learning technologies. The goal to long term is to create a scalable and accurate system capable of providing real-time, street-level data on air pollutants. Where a mobile monitoring system will be implemented using vehicles that are constantly moving around a regional area such as Zacatenco.

2 Related Work

Significant progress has been made worldwide in the monitoring of air quality in terms of PM_{2.5} and PM₁₀ particles. In this section we present, some works that have involved the deployment of monitoring systems in urban areas, to measure air quality and the use of Machine learning models to predict air quality indices based on air quality data. Despite the fact of the interdisciplinary nature of air quality research involves several domains such as atmospheric chemistry, meteorology, and data science for mention some of them.

The use of monitoring stations combined with machine learning models have been studied before, like in [9] where is presented a study of monitoring pollutant levels, specifically PM_{2.5}, PM₁₀, and NO₂, at two locations in Stuttgart is conducted using monitoring station. Machine learning is employed to simulate and predict pollutant concentrations based on meteorological, traffic, and nearby monitoring station data. But the approach does not consider the use of mobile monitoring stations.

While, in [10] The importance of investigating morphological factors that contribute to the spatial distribution of air pollution using mobile monitoring data is emphasized. They assess the nonlinear relationship between the spatial distribution of atmospheric pollutants and the morphological indicators of buildings in a high-density city. They use a mobile station of monitoring but focused on predicting the spatial variability of pollutants and the assessment of six machine learning models where neural networks have a better performance, the testing was made in Shanghai.

Continuing with the machine learning research on air quality, in [11] is shown a study, where is mentioned that there has been a significant increase in the application of ML models in air pollution research. The study is based on a bibliometric analysis of 2962 articles published from 1990 to 2021 and highlights that most publications occurred after 2017. As well as the main research topics related to the application of ML in the chemical characterization of pollutants, short-term forecasting, detection improvement, and emission control optimization. The paper is based on explore the status of ML applications in air pollution research.

For another hand, in [12] research is presented based on spatial-temporal approach combined with machine learning on air quality, they discuss air pollution management and the importance of linking pollution sources to air quality. It mentions that since weather cannot be controlled, air pollution can only be managed by planning and controlling pollution sources. It emphasizes that mathematical models that don't establish a relationship between the source and air quality don't provide guidance on how to regulate air quality. To address this, a machine learning model is used to predict the daily average concentrations of NO₂ and SO₂ in five different locations. Features related to weather and features related to the type and quantity of fossil fuels consumed by power plants are used as proxies for emission rates. Three different models are compared, and Model III is found to be the most accurate, showing significant improvements in pollutant concentration prediction. The paper does not consider the use of mobile stations as our proposal.

Furthermore, in [13] discusses the concern over high air pollution levels in Ho Chi Minh City (HCMC) Vietnam, and the use of machine learning algorithms to forecast air quality in the region. It is mentioned that air pollution levels have exceeded WHO standards, leading to significant impacts on human health and the ecosystem. Is described an effort to develop a global air quality forecasting model that considers multiple parameters such as weather conditions, air quality data, urban spatial information, and temporal components to predict concentrations of NO₂, SO₂, O₃, and CO at hourly intervals. The datasets on air pollution time series were gathered from six air quality monitoring sites from February 2021 to August 2022. It concludes that the global model outperforms previous models that consider only a specific pollutant. Not mobile monitoring is treated, but forecasting is treated.

Forecasting on air quality is addressed in [15] They applied Air Quality Index (AQI) to the city of Visakhapatnam, Andhra Pradesh, India, focusing on 12 contaminants and 10 meteorological parameters from July 2017 to September 2022. Using several machine learning models, including LightGBM, Random Forest, Catboost, Adaboost, and XGBoost. Where Catboost model outperformed other models, they used air quality and meteorological datasets.

The health impacts of air pollution are treated in [14] discussing the alarming increase in air pollution due to industrialization in developing countries and its impact on hospital visits by patients with respiratory diseases, specifically Acute Respiratory Infections (ARI). The study collects data on outpatient hospital visits, air pollution, and meteorological parameters from March 2018 to October 2021. Eight machine learning algorithms were applied to analyze the relationship between daily air pollution and outpatient visits for ARI. The results indicate that, among the eight machine learning models studied, the Random Forest model performed the best. However, it was found that the models did not perform well when considering the lag effect in the dataset of ARI patients.

To conclude, air quality monitoring and prediction, especially regarding PM2.5 and PM10, represent active and evolving research and application areas worldwide. The combination of fixed and mobile sensors, machine learning models, and interdisciplinary research has significantly enhanced our capacity to understand, predict, and address air quality issues in various regions and contexts.

3 Material and Methods

This section presents the system architecture consists of a broker (Azure IoT Hub), a resource for data processing and analytics, which enables working with the stream received by the broker and processing it for storage in a non-relational database using the CosmosDB SQL API it is shown in Fig. 1.

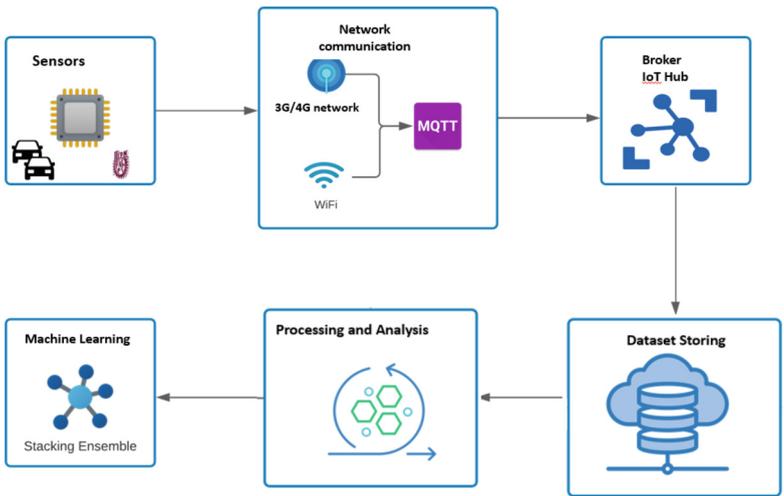


Fig. 1. System Architecture

As Fig. 1 is depicted, each module will be described as follows:

The Sensor Module: This module is responsible for sampling suspended particles (PM10, PM2.5) every hour throughout the day for the stationary station and around 300 samples per day for the mobile station. The sampling frequency for the mobile station was chosen to ensure continuous monitoring and verification of sensor data while in motion. However, this monitoring supervision could be avoided if higher-quality sensors were accessible, such an improvement can be considered for future implementations.

The Network Communications Module: This module is responsible for transmitting the samples taken by the sensor module to the network. The data is processed to determine predictive models that serve as a reference for understanding pollutant behavior.

IoT Broker: It is used to enable secure and bidirectional communication between sensors and cloud applications, enabling secure communication using a set of protocols.

Dataset Storing module: It is responsible for managing and storing the data, from its preprocessing, and for sending it to the data processing module.

Processing and Analysis Module: it is required to prepare data, cleaning and deliver the data to be sent to machine learning module.

Machine Learning module, here is trained, validated the models using the data obtained data module and graphs are generated.

The input data for this project consists of PM10 and PM2.5 levels provided by sensors on the mainboard, a GPS module providing coordinates for hourly measurements, and additional data from the sensors and package transmission time. The route was conducted with a vehicle in the Zacatenco area. Test zones can be stationary or mobile monitoring areas with vehicles that have a constant route throughout the city, as well as public areas where the module for pollutant monitoring can be installed. The monitored data includes PM10 and PM2.5 suspended particles, obtained through the network sensors.

4 Design Network Air Quality

This section is divided into two parts. The first part involves the design, installation, and operation of a sensor network to monitor air quality in the Zacatenco academic facilities of the National Polytechnic Institute in Mexico City. The second part integrates the data obtained from mobile and stationary monitoring and uses a model to predict air quality.

4.1 Design Sensor Network Hardware

This section describes the design of the monitoring stations, which includes the integration of PM2.5/PM10 sensors and a GPS module using two different base boards. It is shown in Fig. 2.

As Fig. 2 shows, one base board is used for the mobile monitoring station, which incorporates the Arduino MKR 1400 GSM board and utilizes the Message Queuing Telemetry Transport (MQTT) protocol for data transmission. The other base board is used for the stationary monitoring station, which incorporates the Arduino MKR 1010 WiFi board and utilizes the MQTT protocol. The PM2.5/PM10 sensor has the following characteristics shown in Table 2.

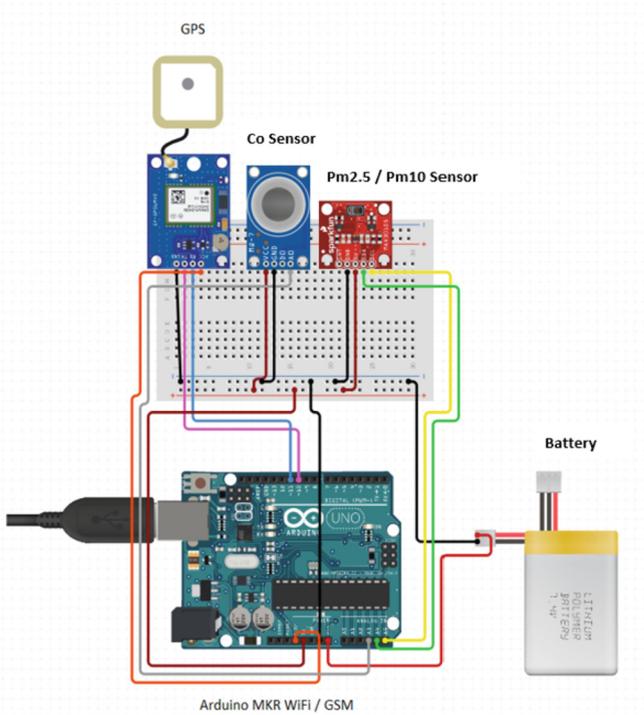


Fig. 2. Hardware of mobile monitoring station.

Table 2. Specifications of suspended Particle Sensor PM_{2,5}/PM₁₀

Specifications	
	Operating Voltage: 4.95 – 5.05 V
	Maximum Electrical Current: 120 mA
	Measurement Particle Diameter: 0.3–1.0, 1.0–2.5, 2.5–10(um)
	Measurement Range for Particulate Matter: 0 – 999 ug/m ³
	Standby Current: ≤200 uA
	Response Time: ≤10 s
	Operating Temperature Range: –20 – 50 C
	Operating Humidity Range: 0 – 99% RH
	Maximum Size: 65 × 42 × 23 (mm)
	MTBF (Mean Time Between Failures):>= 5 years
	Fast Response
	Standard Serial Word Output
	Second-Order Multipoint Calibration Curve
	Minimum Resolution Size of 0.3 microns

4.1.1 Design Stationary Monitoring Station

The stationary monitoring station was designed to take readings of PM 2.5/ PM10 each 15 min per day. The period required by the sensors to obtain accurate samples, which is 5 s. The stationary monitoring station consists of an Arduino board, a PM2.5/PM10 suspended particles module, and a GPS module. The station is shown in Fig. 3.

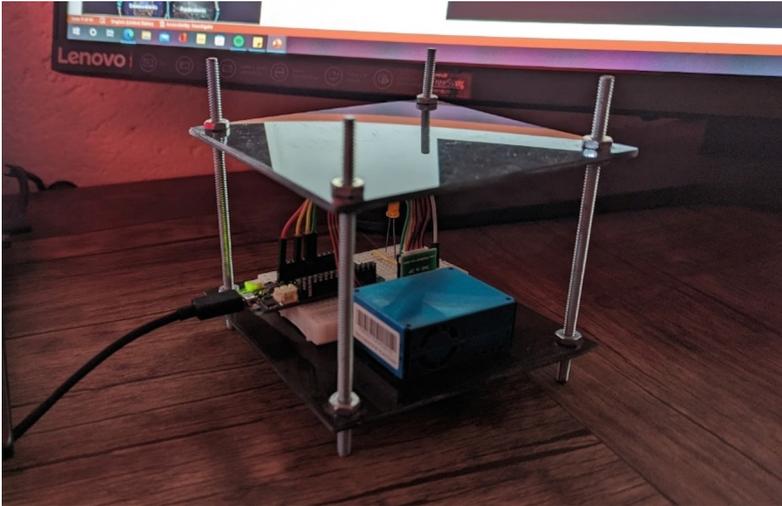


Fig. 3. Stationary station

4.1.2 Design Mobile Monitoring Station

The mobile monitoring station was designed to take readings of PM2.5/PM10 and coordinates every 5 seconds during the routes defined in testing zones. The sensor module was integrated with a module that allows the transmission of the data obtained by the sensors in real-time using either the 4G network or Wi-Fi. It is shown in Fig. 4.

The design of this monitoring station allows it to be transported in an automobile. The average sampling time during each journey was from 10 to 20 min, and it was conducted three times per day, five days a week (Monday to Friday). The integration of the sensor module with the Arduino MKR WiFi 1010 and Arduino MKR GSM 1400 boards enables the periodic transmission of data obtained by the monitoring stations using the available Wi-Fi or GSM/4G network, respectively.



Fig. 4. Mobile Monitoring station

4.2 Analysis of Air Quality Data

Data cleaning was performed to prepare the training and validation data, keeping only the relevant data for the purpose of the machine learning model, which includes the timestamp and the parameters-pm25/value and parameters-pm10/value.

The timestamp column was transformed into a date type, and the execution was performed in two phases: one with PM2.5 suspended particle values and the other with PM10 suspended particle values.

The collected data is stored as JSON documents in the Cosmos DB, a NoSQL database where T-SQL is used for queries and data processing (see Table 3).

Table 3. JSON Document Data Dictionary

Document	Variable Name	Data Type	Description
General	Timestamp	String	Event Identifier with Date and Time
	Stationid	String	Station Identifier
	Parameters	Object	Document with Event Parameters Information
	Latitude	Float	Event Coordinates
	Longitude	Float	Event Coordinates

(continued)

Table 3. (continued)

Document	Variable Name	Data Type	Description
Parameters	Idparameter	String	Parameter Identifier
	Value	Int	Sensor Reading Value
	Unit	Int	Sensor Reading Unit
Obtained from IoT Hub	EventProcessedUtcTime	String	Chain Processing Time
	PartitionId*	Int	Hub Partition ID
	EventEnqueuedUtcTime*	String	Chain Queue Processing Time
	MessageId*	null	Message ID
	CorrelationId*	null	IoT Hub Relationship ID
	ConnectionDeviceId*	String	Device ID
	ConnectionDeviceGenerationId*	String	Connection ID
	EnqueuedTime*	String	Queue Processing Time
	Id*	String	Message ID in Document

A database was provisioned for each of the monitoring stations, the data normalization used the method min-max scaler. This normalization method allows us to transform our features (i.e., the columns of our data [PM2.5/PM10]) within a minimum and maximum value (0 and 1).

The mathematical function for min-max scaler normalization is given by Eq. (1):

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

The selection of model for training was based on the requirement of predicting a single dimension, in this case the suspended particles PM2.5/PM10, then we used a Long Short-Term Memory model (LSTM network and because is scientific literature has been used as in such as weather forecasts and Time Series Prediction, where LSTM is excellent for forecasting future values in time series data, due to its ability to capture long-term patterns in the data.

Long Short-Term Memory (LSTM) method, is a type of Recurrent Neural Network (RNN) that is capable of learning long-term dependencies. It was introduced by Hochreiter and Schmidhuber in 1997 [17]. All recurrent neural networks have a chain-like structure with repeated modules of a neural network. LSTM networks also have a similar structure, but the iteration module has a different structure. Instead of a single layer, it consists of four layers of neural networks that interact in a special way. The

LSTM neural network comprises an input layer, an LSTM layer, a fully connected layer, and an output layer [16], as depicted in Fig. 5.

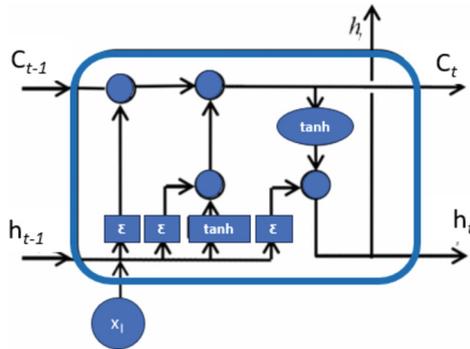


Fig. 5. LSTM representation

The LSTM model works as follows:

- 1) The first step is to decide which information to discard in the node state. This decision is made by the layer called the ‘forget gate layer’. From the output of $h_{(t-1)}$ and x_t , ranges between 0 and 1 are obtained, indicating whether the information is transmitted to the next node or discarded. A value of 1 means the information is needed, while 0 means it should be discarded.
- 2) The next step is to decide which new information to store in the node state. This process is divided into two sub-processes. The first is defined by the sigma function, which determines the values to be updated. The tanh function creates a vector with the new values of the candidates that can be stored in the node state.
- 3) The following step is to perform what was decided in the previous step. We multiply the previous state by f_t , forgetting the previously decided information, and adding the new candidates.

Finally, we decide what to send to the output. This is based on the state of our node. The sigma function determines which part of the node state to send to the output. We then force the values to be within -1 and 1 using the tanh function and multiply them by the output of the sigma function. This ensures that only the data decided in the process is sent to the output.

5 Experiments and Results

The experiments and testing took place over a two-month period spanning from October to November 2022. The configuration involved the installation of three stationary monitoring stations at distinct locations within the Zacatenco region of Mexico City, specifically at IPN Campuses UPIITA, ESCOM, and CIC, respectively. In the case of the mobile monitoring station, a predefined journey sequence was executed. This journey route covered the Zacatenco area and its adjacent avenues. The average sampling

duration for each journey ranged from 10 to 20 min, and this process was repeated three times daily, from Monday to Friday. A visual representation of this route can be found in Fig. 6 on the map.

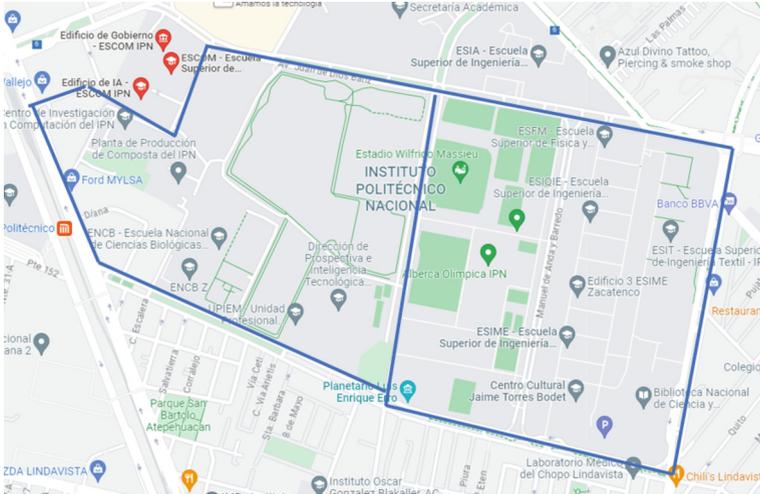


Fig. 6. Geographical area of monitoring in Zacatenco

The two stationary monitoring stations were connected to IPN Wi-Fi network, while the third mobile monitoring station was connected to the GSM/4G network. In Fig. 7 is shown the prototype sensor installation of UPIITA station.

The monitoring process at the UPIITA and ESCOM stations was meticulous, with readings captured at 15-min intervals. This effort resulted in the accumulation of a noteworthy 5,760 data points over the span of two months. This dataset shows the air quality and environmental conditions at these specific locations.

In contrast, the mobile monitoring station employed a more frequent sampling strategy, conducting readings every 5 min. This high-frequency approach led to the collection of a substantial 17,280 samples over the same two-month period. This dataset allows for a more detailed examination of air quality variations in the broader Zacatenco area, offering a comprehensive perspective on the region's environmental dynamics. For a comprehensive overview of the total samples obtained by each monitoring station, including their respective data collection periods, refer to Table 4.



Fig. 7. UPIITA station installation

Table 4. Sample data collection periods

Station ID	Location	Sample Count	Period	Duration
1	ESCOM	17280	September 30 - November 26	2 months
2	UPIITA	5760	October 7 - November 26	2 months
3	CIC	5760	October 26 - November 26	2 month

Figure 8 provides a detailed visual representation of the routes meticulously traced by our mobile monitoring station within the Zacatenco region of Mexico City. These routes were carefully selected to encompass key areas of interest, allowing us to comprehensively assess air quality and environmental conditions across this bustling urban landscape. The map in Fig. 8 not only showcases the routes but also offers essential geographical context, with reference points along the way. By charting these paths, we've been able to gather a wealth of data that contributes significantly to our understanding of air quality dynamics in the Zacatenco area.

This graphical representation allows to visualize the coverage of our mobile monitoring efforts, providing a clear picture of the regions where we have collected critical data for analysis.

As we closely examine the data, we've identified two distinct locations marked on our heat map in red color. One of these points is located at the entrance of campus ESCOM, while the other lies amidst the bustling cross avenues adjacent to the shopping center known as "Torres Lindavista". These specific locations have registered elevated levels of PM 10 and PM 2.5, making them crucial focal points for our analysis.

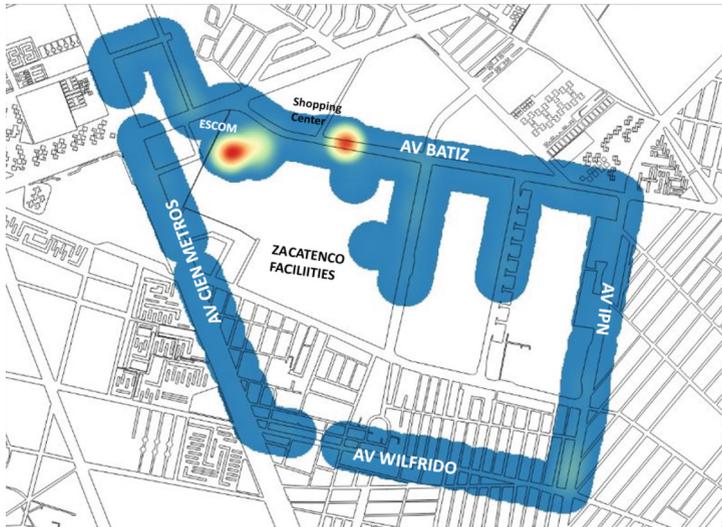


Fig. 8. Geographical area and routes of mobile monitoring in Zacatenco

In our pursuit of building a reliable model, we focused on training and validating it using data from monitoring stations that have accumulated the highest number of samples. As indicated in Table 2, our primary sources of data for model development are the Mobile and stationary monitoring station, which contributed a substantial 17,280 samples, and the second contributor with 5,060 samples.

These stations were strategically chosen due to their data records, ensuring that our model is equipped to provide accurate insights into air quality conditions, especially in areas where historical records have consistently shown heightened PM₁₀ and PM_{2.5} levels. The utilization of these datasets allows to the model predict and address air quality concerns effectively. The Fig. 9 shows the historical behavior obtained from stationary station in UPIITA for PM₁₀ over a 2-month period.

In reference to PM₁₀ levels, it is essential to underscore that the 24-h average remains stable at $75 \mu\text{g}/\text{m}^3$, in accordance with the data provided by the air quality index, accessible at <http://www.aire.cdmx.gob.mx/>. The fact that this value aligns with the recommended limit, as depicted in the accompanying chart, is indeed a positive indication. This implies that, daily, the air quality generally falls within acceptable standards.

Nonetheless, it is imperative to draw attention to a critical point. Towards the end of September, a notably high PM₁₀ value was recorded. This anomaly demands a more comprehensive investigation and heightened attention. It is essential to delve deeper into the circumstances surrounding this spike to better understand the potential causes and implications for air quality management and public health in the region.

Shifting our focus to PM_{2.5}, we have delved into an analysis of its trends, all appears in Fig. 11. This figure serves as a visual aid, enabling us to conduct a more thorough and nuanced examination of PM_{2.5} concentrations and their fluctuations over time.

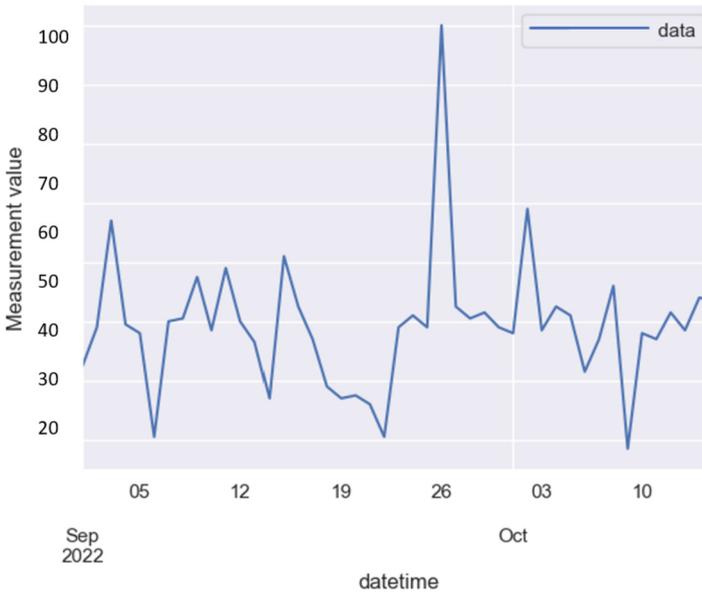


Fig. 9. PM10 Data Historical behavior in UPIITA

Gaining a deep understanding of the dynamics of PM_{2.5} is of paramount importance. These fine particulate matter particles, with a diameter of 2.5 μm or smaller, have a unique ability to linger in the atmosphere and penetrate deeply into our respiratory system when inhaled. As such, comprehending how PM_{2.5} levels evolve and vary over time is crucial, as it provides invaluable insights into the potential impact on air quality and, consequently, public health.

By dissecting the data represented in Fig. 10, we gain a picture of the temporal trends of PM_{2.5}, enabling us to make informed decisions regarding air quality management strategies and health measures.

In the model training process, we have allocated 2/3 of our samples for training the model, enabling the LSTM algorithm to learn complex patterns from a sufficiently dataset.

This constitutes a critical phase in the implementation of machine learning algorithms, such as the LSTM algorithm employed in this study. The data source is from two stations with remarkable total of 17,280 samples, with the second-largest contributor providing 5,060 samples, all collected over a period of two months.

The validation phase becomes essential to assess the model's performance. We make to experiments, one using only the data form stationary station.

The separation of 1/3 of the data for this purpose allows us to verify how the model performs on unseen data and ensures it is not overfitting the training data. In Fig. 11 as can be showed the forecast was not good, the performance was around 70%.

Then, we proceed to integrate and balance all dataset and apply the LSTM therefore, with the model trained and validated, it was applied. The results are illustrated in Fig. 12.

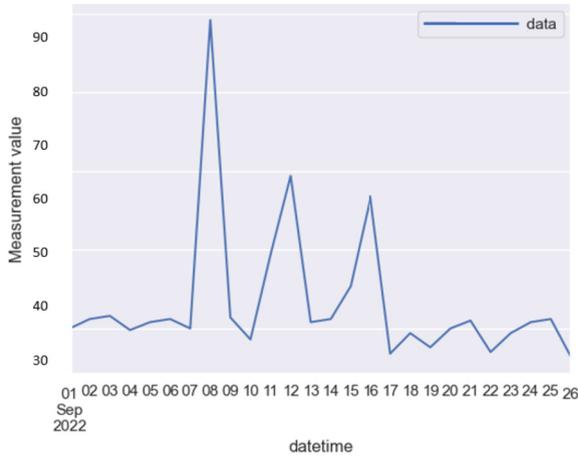


Fig. 10. PM2.5 Data Historical behavior in UPIITA

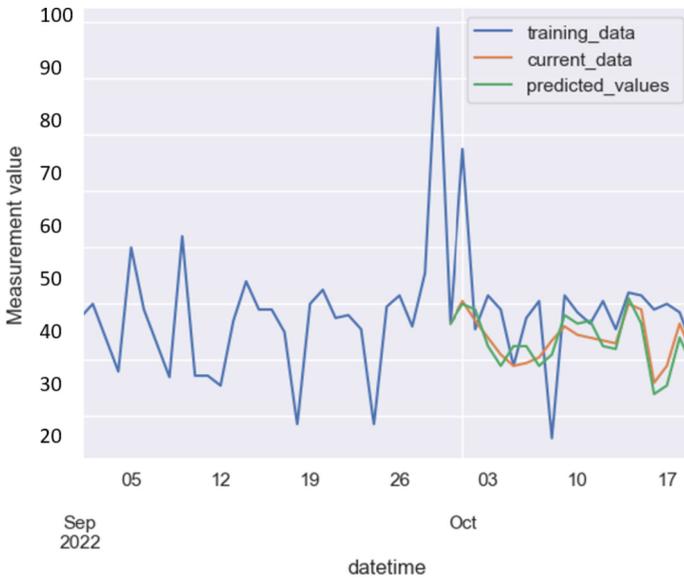


Fig. 11. Forecasting for PM10 only with data from Stationary monitoring station

The performance of LSTM was 88% for the case of PM10 as is shown in graphic. We can conclude the success of the model based on the results obtained.

Finally, the performance metrics of our model for PM10 suspended particles. With a loss of 0.008674 defined by the which indicates the percentage of error obtained by the model.

The monitoring stations fulfill the purpose of taking street-level readings every hour, 24 h a day for the stationary station, and every 5 s for daily routes for the mobile station.

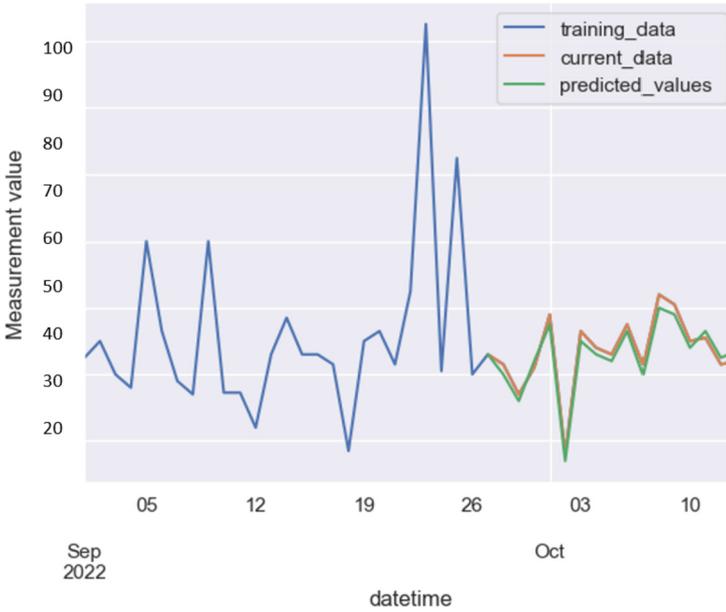


Fig. 12. Forecasting for PM10 with all datasets integrated

6 Conclusions and Future Work

The presented approach successfully demonstrates an effective method for monitoring air quality in regional areas of Mexico City. The hybrid network of mobile and stationary sensors provides a granular street-level perspective on particulate matter behavior. The combination of IoT technology and machine learning techniques proves to be valuable for data processing and generating predictive models for air quality forecasting at street level.

This research contributes to enhancing the understanding of pollution levels in urban environments, supporting informed decision-making to improve public health and well-being. While the model’s overall performance is acceptable, there is room for improvement. Future work should explore additional models and algorithms to enhance accuracy. To gain deeper insights into air quality patterns, long-term data analysis should be considered to identify seasonal variations and correlations with other factors.

The presented approach represents a significant step towards improved air quality monitoring in Mexico City. It highlights the potential for using technology and data-driven methods to address air pollution challenges in urban environments. However, ongoing research and collaboration are needed to refine the system and maximize its impact on public health and environmental well-being.

While in Future work, expansion of Sensor Network, this would allow for data collection across a larger geographical coverage, providing a more comprehensive view of air quality in Mexico City.

Regarding experimenting with different machine learning models and algorithms to improve the accuracy of air quality forecasts. This may include exploring more sophisticated techniques or ensemble methods.

To better understand trends and patterns in air quality, conducting long-term data analysis could be beneficial. This could involve analyzing data over several years to identify seasonal variations, long-term trends, and potential correlations with other factors like weather patterns or traffic. Developing user-friendly data visualization tools and platforms for public outreach Making air quality data easily accessible and understandable to the public can raise awareness and drive behavior change. Continuously validating and calibrating the sensors to ensure data accuracy is crucial. Developing protocols and methodologies for sensor calibration and validation. Overall, the future work could involve refining and expanding the existing air quality monitoring system to enhance its accuracy, coverage, and impact on public health and urban well-being.

Acknowledgments. The authors want to thank God, IPN, COFAA, UPIITA-IPN, and ESCOM-IPN for their support. This work has been partially supported by the IPN SIP-20231645, 20230096, 20231156, CONAHCYT CF-2023-G-1170 and 7051 projects, and MISTI-IPN-MIT Found.

References

1. Guo, P., et al.: Ambient air pollution and markers of fetal growth: a retrospective population-based cohort study of 2.57 million term singleton births in China. *Environ. Int.* **135**, 105410 (2020)
2. Bravo-Alvarez, H., Torres-Jardón, R.: Air pollution levels and trends in the Mexico City metropolitan area. *Urban Air Pollution and Forests: Resources at Risk in the Mexico City Air Basin*, pp. 121–159 (2002)
3. Wang, Z., Delp, W.W., Singer, B.C.: Performance of low-cost indoor air quality monitors for PM_{2.5} and PM₁₀ from residential sources. *Building Environ.* **171**, 106654 (2020)
4. Nelson, P.F.: Environmental issues: emissions, pollution control, assessment, and management. In: *The Coal Handbook*, pp. 31–76. Woodhead Publishing (2023)
5. Dai, X., Shang, W., Liu, J., Xue, M., Wang, C.: Achieving better indoor air quality with IoT systems for future buildings: opportunities and challenges. *Sci. Total Environ.* **895**, 164858 (2023)
6. Hajmohammadi, H., Heydecker, B.: Evaluation of air quality effects of the London ultra-low emission zone by state-space modelling. *Atmos. Pollut. Res.* **13**(8), 101514 (2022)
7. SEDEMA. Calidad del Aire, SEDEMA, 2018. [En línea]. Available: <http://www.data.sedema.cdmx.gob.mx/breatheLife/calidadAire.html>
8. Gobierno de la Ciudad de México. Dirección de Monitoreo Atmosférico, [En línea]. Available: <http://www.aire.cdmx.gob.mx/default.php?opc=%27ZaBhnmI=%27>
9. Samad, A., Garuda, S., Vogt, U., Yang, B.: Air pollution prediction using machine learning techniques – an approach to replace existing monitoring stations with virtual monitoring stations. *Atmos. Environ.* **310**, 119987 (2023)
10. Huang, C., et al.: Effect of urban morphology on air pollution distribution in high-density urban blocks based on mobile monitoring and machine learning. *Build. Environ.* **219**, 109173 (2022)
11. Li, Y., Sha, Z., Tang, A., Goulding, K., Liu, X.: The application of machine learning to air pollution research: a bibliometric analysis. *Ecotoxicol. Environ. Saf.* **257**, 114911 (2023)

12. Alolayan, M.A., Almutairi, A., Aladwani, S.M., Alkhamees, S.: Investigating major sources of air pollution and improving spatiotemporal forecast accuracy using supervised machine learning and a proxy. *J. Eng. Res.* 100126 (2023)
13. Rakholia, R., Le, Q., Ho, B.Q., Vu, K., Carbajo, R.S.: Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City Vietnam. *Environ. Int.* **173**, 107848 (2023)
14. Ravindra, K., et al.: Application of machine learning approaches to predict the impact of ambient air pollution on outpatient visits for acute respiratory infections. *Sci. Total Environ.* **858**, 159509 (2023)
15. Ravindiran, G., Hayder, G., Kanagarathinam, K., Alagumalai, A., Sonne, C.: Air quality prediction by machine learning models: a predictive study on the Indian coastal city of Visakhapatnam. *Chemosphere* **338**, 139518 (2023)
16. Meyal, A.Y., et al.: Automated cloud based long short-term memory neural network based SWE prediction. *Front. Water* **2**, 574917 (2020)
17. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 32 (1997)