Real-Time Machine Learning for Air Quality and Environmental Noise Detection

Conference Paper · December 2020

DOI: 10.1109/BigData50022.2020.9377939

4 authors:

Sayed Khushal Shah
University of Missouri - Kansas City

Zeenat Tariq
University of Missouri - Kansas City

Jeehwan Lee
Myongji University

Yugyung Lee
University of Missouri - Kansas City

Real-Time Machine Learning for Air Quality and Environmental Noise Detection

Sayed Khushal Shah
School of Computing and Engineering
University of Missouri-Kansas City
USA
ssqn7@umsystem.edu

Zeenat Tariq
School of Computing and Engineering
University of Missouri-Kansas City
USA
zt2gc@umsystem.edu

Jeehwan Lee
College of Architecture
Myongji University
Seoul, Korea
jeefanlee@gmail.com

Yugyung Lee
School of Computing and Engineering
University of Missouri-Kansas City
USA
leeyu@umkc.edu

***Abstract*—In metropolitan cities, outdoor air pollution and ambient outdoor noise transmission are significant environmental hazards, degrading indoor environmental quality. Natural ventilation is often a viable option for diluting indoor air pollutants, but transportation noise transmission is a crucial ecological conflict. Therefore, personal control over the concentration of outdoor/indoor air pollutants and noise levels is a significant threshold for natural ventilation availability. This study proposes real-time data detection and notification solutions using sensors and artificial intelligence (AI) to improve indoor air quality and outdoor air quality and outdoor noise transmission. An intelligent real-time detection and notification system was implemented in a distributed computing framework with cloud and edge computing. The objective of this study is placed on three facets: (1) development of sensors and NVIDIA Jetson Nano prototype for air quality and noise level detection, (2) application of machine learning for air quality and noise level prediction and classification, and (3) web interface for real-time monitoring and prediction for air quality and noise level detection. The results showed that the proposed user interface provides building occupants with real-time data of outdoor/indoor Air Quality Index (AQI) and noise levels for the optimized occupant control over Indoor Air Quality (IAQ). The personal control over indoor environmental quality (IEQ) enables occupants to promote natural ventilation behaviors and integrate with the existing building system on optimized IEQ by interacting with AI-based real-time data.**

***Index Terms*—Internet of Things (IoT), Arduino Uno, Jetson Nano, My SQL, Web-interface, Indoor air quality, Indoor environment, Machine Learning**

1. INTRODUCTION

Air or noise pollution can lead to irritation and loss of quality of life that may harm health [1]. The presence of particulate matter, biological molecules, and other harmful materials involve many health risk factors for children, adults, and older people. These risk factors may include different types of cardiovascular diseases, skin allergies, pulmonary diseases [2], etc. Such severe illness can be one the cause of death [3], [4]. Air pollution also harms other living organisms, such as animals and food crops, which can damage the natural environment. Also, an essential environmental exposure that may be related to air pollution is noise pollution. It is not easy to take the fresh air or open the windows due to high street noise or high pollutant concentrations.

On the other hand, the poor air quality of indoor and outdoor can cause numerous adverse health problems, such as sick building-related symptoms. Natural ventilation is often a viable option for diluting indoor air pollutants [5], [6]. However, the outdoor particulate matters (PMs) and outdoor noise levels deter building occupants from natural ventilation availability when the level of outdoor PM2.5 and outdoor traffic noise are higher than 25 g/m3 and 53 dB as a 24-hour average based on the World Health Organization (WHO) [7].

Advanced technologies are required to monitor and detect air quality and noise levels for a healthy life. Internet of Things (IoT) and machine learning play a vital role in improving the quality of living and maintaining the development of cities [8], [9]. To considering diverse environmental requirements such as air pollutants, air temperature, relative humidity, wind speed, and noise levels, personal control is one of the influential factors for user satisfaction and environmental comfort due to its physical and psychological impacts. A higher degree of personal control leads to more satisfaction with IEQ [10]. Therefore, the proposed prototype can provide real-time data with maximum personal control over outdoor and indoor air pollutants and noise levels for occupants' comfort and health. For noise and IAQ monitoring and control, a number of studies utilized wireless sensor networks (WSNs) [11], Arduino and a cloud-based platform with sensors [12], a ZigBee network [13], an IoT-based system [14], [15], and a mobile device with read quick response (QR) code for real-time IAQ monitoring system [16].

Many researchers have leveraged vast amounts of streaming data analytics with the capabilities of cloud infrastructures

978-1-7281-6251-5/20/$31.00 ©2020 IEEE

and services [17]. However, there are intense demands for edge computing [18] that supports rapid in-situ analytic in smaller scale platforms like IoT devices or mobile devices. Our research is inspired to overcome the challenge leveraged real-time machine learning with the internet of things (IoT) on edge. Though limited to real-time environmental monitoring and detection, this prototype system is the first in a new wave of IoT-based real-time monitoring and detection of ecological AQI and noise level.

The main contributions of this paper can be summarized as follows: First, we proposed an IoT-based real-time monitoring and detection system that is composed of sensors, Arduino Uno, and NVIDIA Jetson Nano to detect air quality and level of noise. Second, we designed a machine learning-based real-time prediction model, which aims to solve air quality and noise problems in the future. Third, we create a web interface for real-time monitoring and reporting for air quality and noise level detection. NVIDIA Jetson Nano prototype equipped with sensors, web interface, and machine learning enables building occupants to access real-time data for optimized personal control of IEQ.

2. Related Work

Han et al. [19] forecasted the Air Quality Index (AQI) for cities in the United States using machine learning. They analyzed data and developed a machine learning model using Weka for forecasting two primary pollutants in the air, including nitrogen dioxide (NO2) and ground-level ozone (O3). They reported the Support Vector Machine outperformed other machine learning algorithms in terms of the MAE and RMSE values. Chen et al. [20] studied the air quality measures of 16 big cities in China using physical sensors. They identified the chemicals, which affected air quality and discovered their relationship. They predicted air quality index with 22 common significant factors using a PEK based machine learning, i.e., ensemble artificial neural networks. The factors analysis was conducted using a PMI based IVS method for air quality. The result showed that PM25, PM10, and SO2 are the common factors affecting the air quality index, and they are strongly related to each other.

Xi et al. [21] performed urban air pollution detection. Due to the WRF-Chem model not updated regularly, it provides a better prediction model. They considered features from the WRF-Chem model and designed a comprehensive framework. The new structure helped in improving the prediction at a higher level than the WRF-Chem prediction model found 74 cities in China. They used support vector machines, decision trees, random forest, and gradient boosting. As a result, the accuracy was between 70 - 85% for the top 10 cities, while 40 - 53% for the lowest ten cities. they concluded that they obtained better accuracy with more features. Raj et al. [22], a comprehensive air quality study, was conducted in the smart cities of Denmark using low-cost sensors. Machine learning algorithms were used for binary and multiclass classification with artificial neural networks, support vector machines, and multiclass support vector machines. They reported SVM performed better than ANN. Furthermore, the polynomial SVM performed the best of all, with approximately 95% for anomaly detection.

Kumar et al. [23] proposed a hardware-based approach for measuring the air quality index. The hardware is composed of a Raspberry Pi connected to Node-Red for visualization of sensors values and uploading the data to the IBM Bluemix Cloud server. They tested performed on the hardware deployed in Delhi, India, and produced a more visually appealing result from low costs sensors as compared to weather station results. Zhen et al. [24] developed a sensor network based hardware system for monitoring of air quality index. Sensors are connected through a low-cost sensor network to communicate with the server hosted in the cloud. The real air quality monitoring system was compared to the local official air quality station, and an application was developed for end-users based filtering for the air quality.

Marin et al. proposed an efficient alternative to expensive local infrastructures [25]. They considered deploying sensors to numerous areas in cities for air quality monitoring. They observed that an adequate amount of data was collected from sensors in multiple fields, and the data were hosted on a cloud server. However, the values were less reliable due to the lack of an adequate calibration of sensors. They concluded that the effort to transform existing air quality monitoring from local infrastructures to sensor/cloud is not trivial. Still, the sensor/cloud environment is an economical alternative in flexibility considerations.

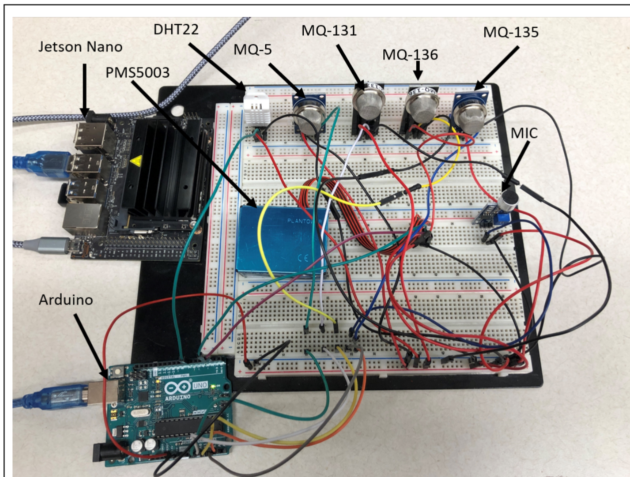

Fig. 1: Real-time Air Quality and Noise Detection System

3. System Design and Methodology

A. System Design

We have proposed to develop a real-time machine learning system using a set of low-cost gas, pollution, and noise sensors connected to an Arduino microcontroller and NVIDIA Jetson Nano. This system is a client-server architecture with two major components, including IoT component and Cloud computing component. The client component is based on an

IoT based edge computing platform, including Sensors, Arduino microcontroller, and NVIDIA Jetson Nano. The server component is based on a cloud computing platform.

The IoT component is connected to the back end using WIFI dongle that enables more flexible and real-time detection in a mobile platform. The training for machine learning would happen in the cloud while the real-time detection would conduct in NVIDIA Jetson Nano at the edge. When the Internet is not available, the data collected from sensors will be saved locally for future use. Our proposed prototype is shown in Figure 1.

We show that the proposed IoT sensing design can offer a low cost for easy integration of multiple sensors to maximize various detection, including gas, pollution, and noise level while maintaining the detectors' reliability at desired levels under different environmental conditions. The overall dimensions of 13 x 11 inches in width and height have been decreased. Also, there are advantages to its low cost and convenience of use. The total cost of the hardware, including sensors, Arduino, and NVIDIA Jetson Nano, is approximately $350. The system was implemented based on open source projects for a web interface for IoT and real-time machine learning.

The hardware components used in our system are described:

Sensing Unit: The sensing unit consists of sensors required for air quality conditions (AQI) and noise level detection. Considering the size, cost, and reliability of an initial prototype, we have selected the MQ gas sensors [26] for the detection of chemical gases in the air and an electric microphone for the discovery of the decibel level of noise in the surroundings. More specifically, Table I shows the MQ-5 sensor was used for carbon monoxide (CO), MQ-131 for ground-level ozone gas detection, MQ-136 for the detection of SO2, MQ-135 gas sensor for nitrogen dioxide and PMS5003 sensor for detection of particulate matters (PM10 and PM2.5). All of the sensors are attached to Arduino, an open-source electronic prototyping platform enabling users to create interactive electronic objects. For the communication between sensors and Arduino, analog channels were used. Table I shows the ranges and units for air quality and the noise data recording device and frequency. The audio signal was recorded for 5 seconds in wav format during different times of the day. The Arduino sends the analog data collected from the sensors to the NVIDIA Jetson nano for digital processing.

Arduino Uno: Arduino UNO is a microcomputer which is based Microchip ATmega328P microcontroller [27]. The board in Arduino UNO is an open-source board composed of analog and digital pinouts that serve as specific purpose I/O ports and need a low power of 9V. Arduino provides an integrated development environment (IDE) to computer programmers for simple application development with input from sensors via digital and analog pins. For enhancing the capability, the Arduino boards can be working with other embedded systems, i.e., Raspberry and NVIDIA Jetson Nano. In our project, the data collected through serial communication with the Arduino board's sensors are processed in a more robust edge device, NVIDIA Jetson Nano, for real-time machine learning.

Jetson Nano: NVIDIA Jetson Nano [28] is a small, powerful AI computer that includes the popular ML frameworks like TensorFlow, Keras, PyTorch, and supports to run AI applications with high performance. The NVIDIA Jetson Nano is compact, low-power, and low-cost and follows the ARM architecture for the CPU chip. The board has a high-quality GPU video card, which supports high-performance computing. Competing with Raspberry Pi, the Nano has GPIO (General Purpose Input/Output) pins. The Jetson Nano can run a variety of Linux-based systems, but it has massive potential in edge computing, specifically supporting real-time Machine Learning applications. GPU-based architecture can solve Machine Learning tasks efficiently compared to Raspberry Pi or even CPU-based modern computers. The Jetson Nano is the main component in our system that saves the analog data from Arduino in the MySQL database. The collected data would be used for visualization through our web interface and refinement of the machine learning models. The MySQL server, Web server, and the Machine Learning server are all hosted in the NVIDIA Jetson Nano for performing the machine learning of tasks in real-time.

MySQL: MySQL [29] is an open-source relational database management system that supports a structured query language SQL. MySQL is one of the leading and most commonly used platforms for adding, accessing, storing different process types of data into physical files. The MySQL applications, including content management system websites, mobile phone applications with database access, can lead to large scale organizational level database management. In our project, the MySQL database is used to store real-time data transferred from Arduino initially from the sensing unit, and further is used for Web interface visualization and machine learning processing.

Web Interface: The web interface is designed in PHP and HTML5.0. Due to the development features, it has all advanced features to update the real-time air quality index and environmental noise level for different cities in South Korea. The HTML5 based Web application using standards-based APIs automatically fits any screen size for supporting real-time Machine Learning applications. The web applications can be deployed in Android and iOS mobile apps through the Ionic framework that is a complete open-source SDK for hybrid mobile app development. The development of mobile apps is considered for future work.

Machine Learning Server: The machine learning server runs on Jetson Nano in our system, and it mainly supports real-time predictions on the data captured from the sensors. The prediction results have been presented on the website. The projections have been made two different modes: first, the real-time prediction has been made for a specific time and place and second, a periodic prediction such as every 24 hours using data stored in the MySQL server, and the results are shown for end-users with the predicted air quality and noise level of a specific city for the next 24 hours.

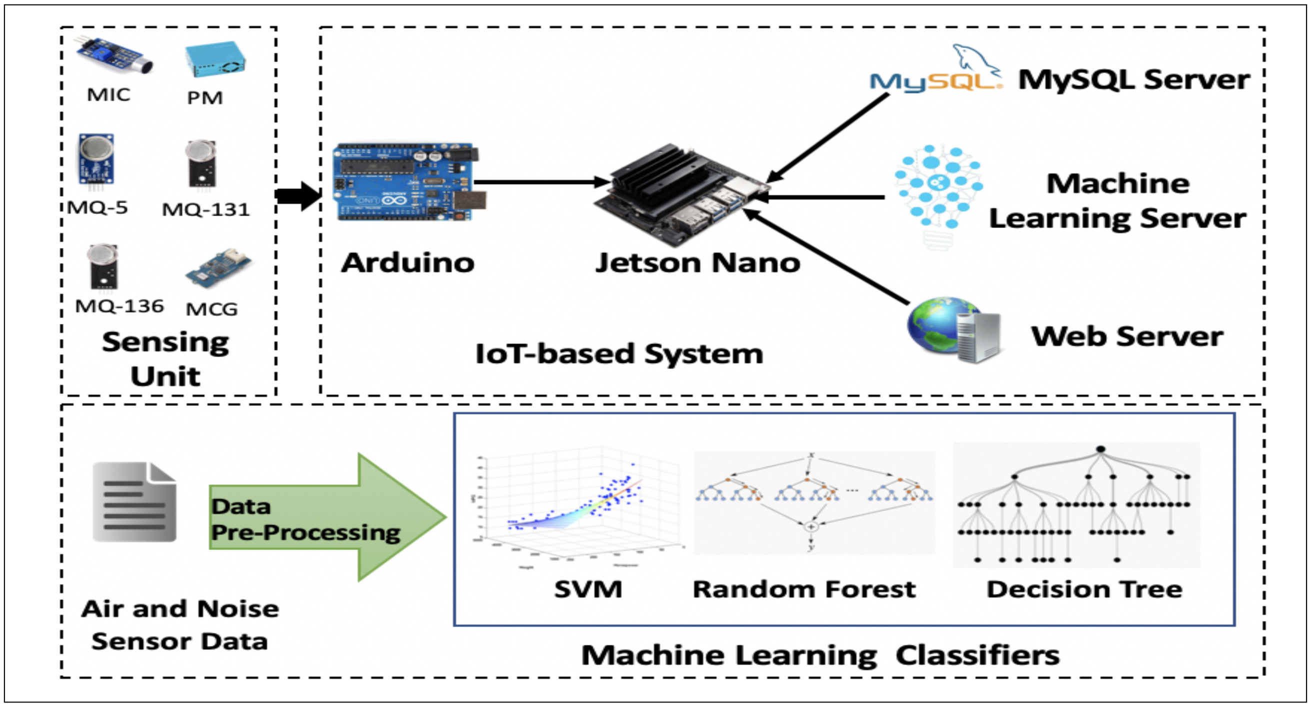

Fig. 2: Prototype Diagram for Air Quality/Noise Prediction

TABLE I: Hardware Components Used for Air Quality and Noise Level Detection

Sensors/IoT	Devices used for AQI Detection	Configuration
MQ-5	Natural gas, LPG	Sensing Resistance: 10 KΩ–60$K\Omega$;
MQ-131	Ground level Ozone	Concentration Scope: 200 - 10000ppm;
MQ-135	CO, Ammonia, Benzene, Alcohol, smoke	Operating voltage: 5V;
MQ-136	Hydrogen Sulfide gas (SO2)	Heater resistance: 33Ω5%
PMS5003	Detection of particulate matters (PM10 and PM2.5)	
DHT22	Digital temperature/humidity sensor to measure the surrounding air	Humidity: 0-100% RH; Temperature: -40°-80°C
Mic	USB microphone	Samsung Galaxy S6
Arduino UNO	Open-source electronic prototyping platform enabling users to create interactive electronic objects	Microcontroller ATmega328P: 8 bit AVR family microcontroller; Input Voltage Limits: 6-20V; Analog Input Pins: 6 (A0 – A5); Digital I/O Pins: 14; DC Current on I/O Pins: 40 mA
Jetson Nano	NVIDIA's embedded system-on-module and developer kit. Used for real-time machine learning (ML) by deploying ML models to be run on the Edge device.	microSD card (16GB UHS-1 minimum), USB keyboard and mouse, Computer display (either HDMI or DP), Micro-USB power supply (5V2A)

B. Machine Learning Algorithms

We have applied three machine learning algorithms for the detection of air quality and noise levels. These are decision tree, random forest, and support vector machine.

Decision Tree: The decision tree algorithm is a supervised learning algorithm used for classification and regression. The algorithm aims to generate a model that learns simple decision rules concluded from the data features and predicted the value of the class label. A decision tree is composed of (N, E) so that an internal node N_i represents a test on an attribute, an edge E represents the outcome of the test, and a leaf node N_l represents a class label. The classification rules can be generated from the paths from the root to the leaf.

The construction of the decision tree is based on the information gain of T, which an attribute a_i with the maximum difference between the apriori entropy $\mathrm{H}(T)$ and the conditional entropy $\mathrm{H}(T|a)$ is selected. The nodes expand until all leaves are pure.

Random Forest: Random forest is an extension of the decision tree. Random forests are an ensemble learning by constructing a multitude of decision trees for classification and regression. It produces the classes for classification and the mean prediction of the individual trees for regression. Random forests focused on resolving decision trees' overfitting problems.

Support Vector Machine: Support vector machine is one of the most popular machine learning algorithms. The SVM algorithm builds a model that aims to maximize the margins of classes, i.e., to find a maximum distance between data points of classes in N-dimensional feature space. In the SVM algorithm, the loss function is designed to measure the margin between the data points and the hyperplane.

For given a training dataset of n points such as $(\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)$, y_i indicating the class to which the point $\vec{x}_i$ belongs, $\vec{w}$ is the normal vector to the hyperplane, and $\frac{b}{\|\vec{w}\|}$ determines the offset of the hyperplane from the origin along the normal vector $\vec{w}$. SVM can be formalized with an objective function as in EQ 1.

$$\min_{\frac{1}{n}} \sum_{i=1}^{n} \zeta_i + \lambda \|w\|^2 \tag{1}$$

where $\zeta_i = \max(0, 1 - y_i(w \cdot x_i - b))$, and subject to $y_i(w \cdot x_i - b) \geq 1 - \zeta_i,$ and $\zeta_i \geq 0$, for all i.

4. Results and Discussion

We first discuss the dataset, experimental setup, data analytic, and workflow setup that have been designed for the experiments. Also, we present the results and the web interface for the real-time machine learning application.

A. Dataset and Preprocessing

For machine learning, we used two air quality and environmental noise datasets available for public research purposes. These datasets are "Seoul air quality" [30] and "Seoul national university accommodation noise data" [31], respectively. The air quality dataset is collected and calibrated according to the international standard on air quality for ambient air, one component of which is the mean PM2.5 level over 24 hours for 2014 to 2020. The parameters include the gas data, including CO, NO2, O3, SO2, PM25, PM10.

The noise dataset is collected for the date, distance, floor, XYZ-position, types of sounds, and floor level. The noise dataset (SNU-B36-50) are recordings of NBF in Building 36 at Seoul National University (SNU) with a single built-in microphone of a smartphone (Samsung Galaxy S6). The audio clips are in a wav format and were extracted to the audio features for the position and type. The audio clips include metadata such as the distance from the source, area, and location of the source.

These two datasets were normalized. For example, in the air quality dataset, the data type of CO was in "int64," while most of the others were in "float64". Date data type was in the string while it has been in "Data/Time," these data have been converted to has a uniform data type in data and time data type. Null values were removed from the data. Similarly, the diverse data types of noise datasets are merged. There were still some values that could not be normalized or were NULL values (i.e., no data were received from the sensors).

There may be errors while providing an output of by the sensors every millisecond. It is due to the nature of the sensors during the continuous flow of sensor data. There is a certain standard defined by the sensors developing engineers, which has to follow to get accurate results. Calibration is a practice to improve the performance of the sensors. Table I shows the description and configuration of the sensors and devices we used in this study. For this purpose, we calibrate our gas based on the concentration of a gas in part per million (ppm) by using the resistance ratio of the sensor (RS/R0) where RS is the resistance of the sensor that changes according to the level of gas, and R0 is the resistance of the sensor at a known level without the presence of other gases, or in the fresh air [32].

B. Experimental Setup

The design of our system is well suited for real-time monitoring and prediction of air quality as well as noise level. It is also serving as a working model for real-time machine learning applications at the edge. The experimental setup is designed and constructed for the empirical evaluation for obtaining a reliable prediction model dealing with real-time data from sensors and IoT devices.

The proposed system is composed of essential components, including environmental sensors, Arduino, and NVIDIA Jetson Nano. The circuit setup is shown in Figure 1. For detection of the gases, different types of gas sensors used: MQ-5 for CO (carbon mono oxide), MQ-135 gas sensor for NO2 (Nitrogen Dioxide), MQ136 for SO2 (Sulphur Dioxide), MQ-131 for O3 (Ground-level ozone), PMS5003 for PM25/PM10 (Particulate Matters). For the detection of the noise intensity level in a specific region, we used electret MIC with integrated amplifier LM393. The type of MIC we used is designed for detecting sensitive noise, i.e., background noise or direct and indirect noise.

C. Web Interface for Real-Time Prediction

The web server is capable of real-time monitoring and prediction application by using data from sensors. As the sensors are designed to transmit data only via the analog channel, it was required to use Arduino. Calibration was also applied for getting more accurate readings from the sensors. Furthermore, we have equipped with the MySQL server as edge-based storage, and the data have been utilized for real-time monitoring and automatically prediction via the web interface. As Arduino itself was not sufficient enough to achieve computing-intensive tasks like real-time machine learning, the NVIDIA Jetson Nano was used for more intensive computing with the data from Arduino. The proposed system is capable of reading, monitoring, and saving data from multiple heterogeneous sensors, real-time detection and prediction, and visualizing the real-time data on the web interface.

The real-time sensor data are visualized in a web-based server, as shown in Figure 3. The visualization shows the current noise intensity-level, temperature, humidity, carbon monoxide (CO), nitrogen dioxide (NO2), sulfur dioxide (SO2), particulate matter (PM10/2.5) and ground-level ozone (O3). We also show the overall air quality for that specific time (window for 24 hours).

Figure 4 shows the prediction for the 24 hours in Seoul that made based on real data from the past, and it is visualized on a real map. By applying Machine Learning, it shows the forecast for 24 hours on the Seoul map. According to the standards defined for AQI and noise level, we have used our ML algorithm to predict the specific time for the real-time prediction visualizations. The standard level of the threshold for noise and AQI is also shown on the web interface.

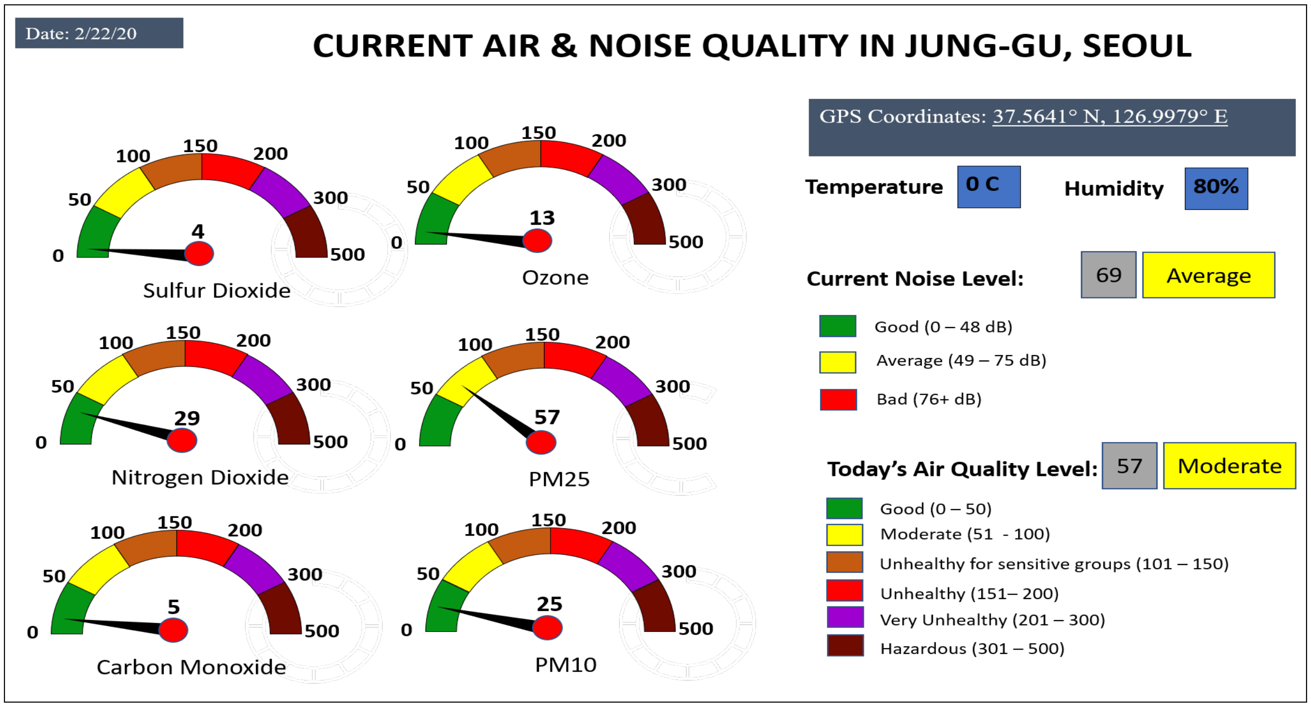

Fig. 3: Web Interface for AQI/Noise Level Real-Time Detection

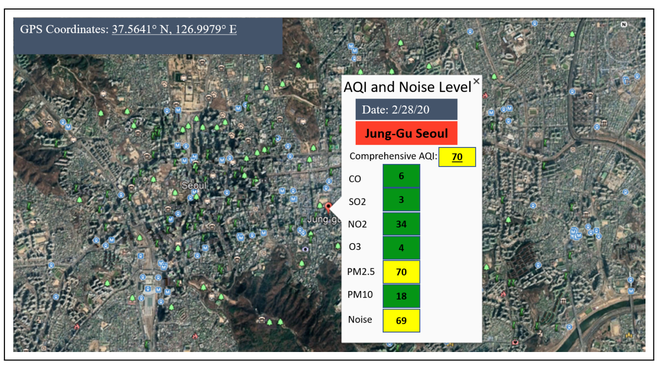

Fig. 4: Map-based Visualization for AQI/Noise Prediction

D. Results and Evaluation

The workflow of our approach is shown in Figure 2. The sensors are capable of recording different air quality gases and also the current noise intensity-level of the surroundings, as explained previously. The results are regarding two main machine learning tasks, such as classification and prediction. It is essential to train the machine learning classifier in such a way that we get real results for the real-time sensor data taken every hour from live sensors. For this purpose, we have chosen three algorithms for classification. These algorithms are decision tree, random forest, and the Support Vector Machine (SVM). SVM has shown better results because of its superior among the supervised machine learning algorithms used for classification and regression problems. The SVM classifier uses a technique called the kernel trick to transform the data, and then based on these transformations, and it finds an optimal boundary between the possible outputs.

Initially, we analyzed the air quality sensor data, which are available publicly from the state department for the Air Quality Index (AQI). When we applied the Support Vector Machine to this dataset, we obtained an accuracy of 75% for daily air quality prediction. The primary reasons for that poor performance were related to data quality issues, e.g., inconsistent formats and missing values. We have fixed a leading cause of low accuracy by adjusting the data types and filtering null values.

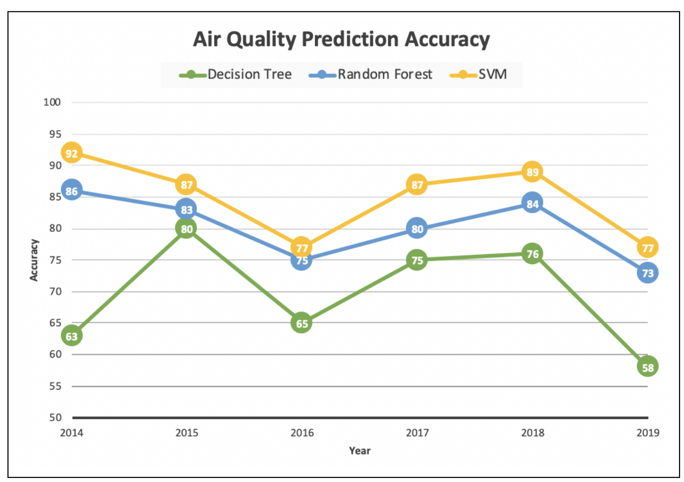

Fig. 5: Comparison of Air Quality Prediction Accuracy

The overall accuracy for the three classifiers for the Seoul air quality dataset [30] is shown in Table II. After removing

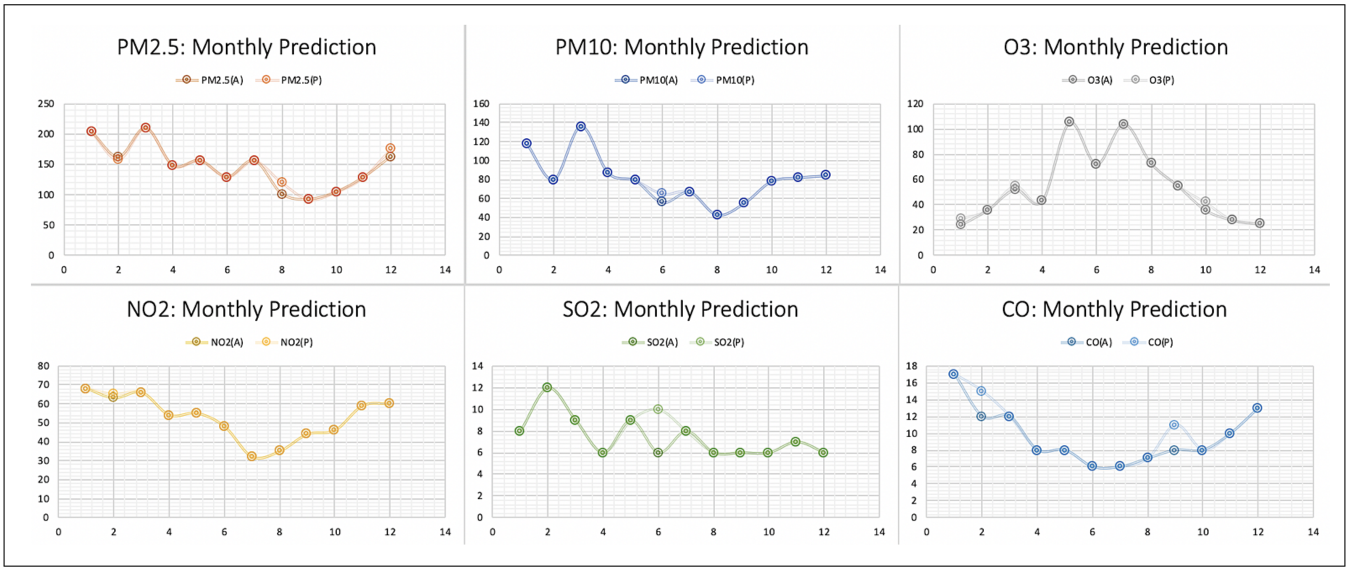

Fig. 6: Air Quality Sensor Prediction in 2019

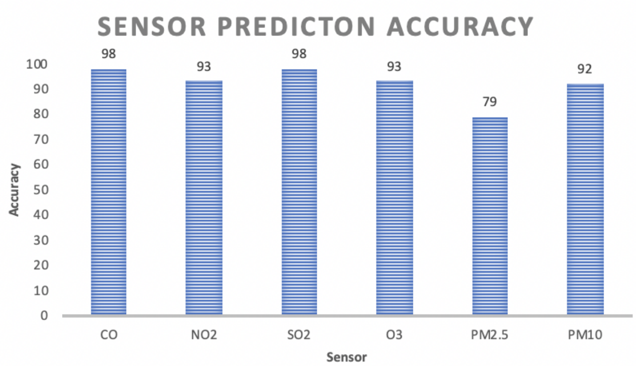

Fig. 7: Sensor Prediction Accuracy for Air Quality

the null values and normalizing the data, the overall accuracy obtained for air quality through SVM was 95%. In contrast, the accuracy obtained from the decision tree and the random forest was 76% and 82%, respectively. SVM shows the highest accuracy compared to the decision tree and random forest.

After the air quality data, we have also used noise data based on Seoul National University's accommodation noise data [31] to detect the noise level of a specific area. The SVM algorithm yields an average classification accuracy of 98% based on data normalization with a split into 70% training and 30% testing data. The accuracy of each model is shown in Table II.

TABLE II: Air Quality & Noise Level Detection Accuracy

Model	Air Quality Accuracy	Noise Level Accuracy
Decision Tree	76%	85%
Random Forest	82%	91%
SVM	**95%**	**98%**

We have also applied the time-based approach and evaluated the data for each year separately for six years of 2014 - 2019, where SVM has shown the best performance for the year of 2014.

Figure 5 shows the comparison of accuracy for air quality index. The data used is for Seoul's official air quality index data. The figure shows the comparison of SVM, random forest, and decision tree. The highest accuracy is shown by SVM followed by random forest and decision tree. As the year progressed, the accuracy drops at some points which are due to inconsistent data. This issue can be overcome using data normalization techniques that can be addressed in future research. Random forest performs faster than the other two models. The SVM classifier is a little slower but reports better accuracy due to multi-class binary classification in nature. The prediction accuracy obtained for SVM was 92%, while the accuracy obtained from the random forest and decision tree was 63% and 86%, respectively.

We have also evaluated our model based on the time-based approach. Specifically, we used the Seoul air quality dataset that reported 24 hours' air quality data for five years in Seoul. Figure 6 shows the year 2019 for sensor performance for the six different sensors (PM2.5, PM10, O3, NO2, SO2, CO) in the air quality prediction. By applying the SVM model to obtain the accuracy for each sensor separately, we have achieved the highest accuracy from the month wise for the year 2019.

The original data was used to verify the accuracy of the predicted value in the real-time end-user prediction. Figure 7 shows the original and predicted values for the six different sensors using a time-based approach for the year 2019. It can be seen from the figure that there is very little difference in predicted values compared to original values. CO and SO2 sensors show the highest accuracy of 98%, while PM2.5 sensor shows the lowest accuracy of 79%. From these experiments, we can conclude that our prediction is pretty accurate.

Some limitations need to be considered: computational power, battery capacity, and communication capacities of the sensors and IoT devices, reliability and security of sensor devices, the robustness of the IoT devices to variations in environmental issues, and the location of the system will be deployed. Also, the lack of users' access to the internet is one of the major concerns.

5. Conclusion

In this paper, we have developed a real-time IoT-based system that enables real-time machine learning of air quality and environmental noise prediction in the surrounding area. We focused on its low cost, portability, and small size for the targeting deployment for the system design with sensors and GPU edge devices. From the real-time machine learning experiments, the SVM model outperformed all other models for real-time prediction. The web interface included a map view for visualization of the predicted results for end-users. We will extend the proposed solution with deep learning technologies to improve the system's performance and reliability.

In the future, we can work on improving the IoT system design and work on the power back up to the system. Applying deep learning techniques could be considered as one of the key features to improve classification accuracy.

References

[1] A. Gupta, A. Gupta, K. Jain, and S. Gupta, "Noise pollution and impact on children health," *The Indian Journal of Pediatrics*, vol. 85, no. 4, pp. 300–306, 2018.

[2] Z. Tariq, S. K. Shah, and Y. Lee, "Lung disease classification using deep convolutional neural network," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2019, pp. 732–735.

[3] R. D. Brook, S. Rajagopalan, C. A. Pope III, J. R. Brook, A. Bhatnagar, A. V. Diez-Roux, F. Holguin, Y. Hong, R. V. Luepker, M. A. Mittleman *et al.*, "Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the american heart association," *Circulation*, vol. 121, no. 21, pp. 2331–2378, 2010.

[4] K.-H. Kim, E. Kabir, and S. Kabir, "A review on the human health impact of airborne particulate matter," *Environment international*, vol. 74, pp. 136–143, 2015.

[5] J. Lee, "Acoustical perceptions of building occupants on indoor environmental quality in naturally-ventilated building façades," *Journal of Acoustics*, vol. 4, no. 3, 2019.

[6] G. Brager and L. Baker, "Occupant satisfaction in mixed-mode buildings," *Building Research & Information*, vol. 37, no. 4, pp. 369–380, 2009.

[7] W. H. Organization *et al.*, "Air quality guidelines for europe," 2000.

[8] S. K. Shah, Z. Tariq, and Y. Lee, "Audio iot analytics for home automation safety," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 5181–5186.

[9] Z. Tariq, S. K. Shah, and Y. Lee, "Speech emotion detection using iot based deep learning for health care," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 4191–4196.

[10] M. Kwon, H. Remøy, A. van den Dobbelsteen, and U. Knaack, "Personal control and environmental user satisfaction in office buildings: Results of case studies in the netherlands," *Building and Environment*, vol. 149, pp. 428–435, 2019.

[11] A. Lambebo and S. Haghani, "A wireless sensor network for environmental monitoring of greenhouse gases," in *Proceedings of the ASEE 2014 Zone I Conference. University of Bridgeport, Bridgpeort, CT*, 2014.

[12] G. B. Fioccola, R. Sommese, I. Tufano, R. Canonico, and G. Ventre, "Polluino: An efficient cloud-based management of iot devices for air quality monitoring," in *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*. IEEE, 2016, pp. 1–6.

[13] K. F. Tsang, H. R. Chi, L. Fu, L. Pan, and H. F. Chan, "Energy-saving iaq monitoring zigbee network using vikor decision making method," in *2016 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, 2016, pp. 2004–2009.

[14] M. Firdhous, B. Sudantha, and P. Karunaratne, "Iot enabled proactive indoor air quality monitoring system for sustainable health management," in *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*. IEEE, 2017, pp. 216–221.

[15] S. K. Shah, Z. Tariq, and Y. Lee, "Iot based urban noise monitoring in deep learning using historical reports," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 4179–4184.

[16] R. C. Chen, H. Y. Guo, M. P. Lin, and H. S. Lin, "The carbon dioxide concentration detection using mobile phones combine bluetooth and qr code," in *2014 IEEE 6th International Conference on Awareness Science and Technology (iCAST)*. IEEE, 2014, pp. 1–6.

[17] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. Mccauley, M. Franklin, S. Shenker, and I. Stoica, "Fast and interactive analytics over hadoop data with spark," *Usenix Login*, vol. 37, no. 4, pp. 45–51, 2012.

[18] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, M. Yunsheng, S. Chen, and P. Hou, "A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure," *IEEE Transactions on Services Computing*, vol. 11, no. 2, pp. 249–261, 2017.

[19] W. Han, S. Zhai, J. Guo, S. Lee, K. Chang, and G. Zhang, "Analysis of no2 and o3 air quality indices and forecasting using machine learning models," *SAR Journal*, vol. 1, no. 3, pp. 107–114, 2018.

[20] S. Chen, G. Kan, J. Li, K. Liang, and Y. Hong, "Investigating china's urban air quality using big data, information theory, and machine learning." *Polish Journal of Environmental Studies*, vol. 27, no. 2, 2018.

[21] X. Xi, Z. Wei, R. Xiaoguang, W. Yijie, B. Xinxin, Y. Wenjun, and D. Jin, "A comprehensive evaluation of air pollution prediction improvement by a machine learning method," in *2015 IEEE International Conference on Service Operations And Logistics, And Informatics (SOLI)*. IEEE, 2015, pp. 176–181.

[22] R. Jain and H. Shah, "An anomaly detection in smart cities modeled as wireless sensor network," in *2016 International Conference on Signal and Information Processing (IConSIP)*. IEEE, 2016, pp. 1–5.

[23] S. Kumar and A. Jasuja, "Air quality monitoring system based on iot using raspberry pi," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*. IEEE, 2017, pp. 1341–1346.

[24] K. Zheng, S. Zhao, Z. Yang, X. Xiong, and W. Xiang, "Design and implementation of lpwa-based air quality monitoring system," *IEEE Access*, vol. 4, pp. 3238–3245, 2016.

[25] M. B. Marinov, I. Topalov, E. Gieva, and G. Nikolov, "Air quality monitoring in urban environments," in *2016 39th International Spring Seminar on Electronics Technology (ISSE)*. IEEE, 2016, pp. 443–448.

[26] C.-y. Chiu and Z. Zhang, "The air quality evaluation based on gas sensor array," in *2017 China Semiconductor Technology International Conference (CSTIC)*. IEEE, 2017, pp. 1–5.

[27] S. A. Arduino, "Arduino," *Arduino LLC*, 2015.

[28] O. Debauche, S. Mahmoudi, P. Manneback, and A. Assila, "Fog iot for health: A new architecture for patients and elderly monitoring." *Procedia Computer Science*, vol. 160, pp. 289–297, 2019.

[29] A. W. West and S. Prettyman, *Practical PHP 7, MySQL 8, and MariaDB Website Databases: A Simplified Approach to Developing Database-Driven Websites*. Apress, 2018.

[30] "Air Quality Index 2014 - 2020, howpublished = https://aqicn.org/city/seoul/, note = Accessed: 2020-02-10."

[31] H. Choi, S. Lee, H. Yang, and W. Seong, "Classification of noise between floors in a building using pre-trained deep convolutional neural networks," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2018, pp. 535–539.

[32] Jaycon systems "understanding a gas sensor. [Online]. Available: https://jayconsystems.com/blog/understanding-a-gas-sensor