IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, August 2024

# Air Quality Prediction System for Smart Cities

**Sowmiya T[1], Prof. D. Shanthi Chelliah[2], Mrs. D. Jagadeeshwari[3], Mr. V. Praveen[4]**
II$^{nd}$ Year ME-Applied Electronics[1]
HOD/ECE, Department of Electronics and Communication[2]
AP/ECE, Department of Electronics and Communication[3]
Sri Muthukumaran Institute of Technology, Chennai, Tamil Nadu, India

**Abstract:** *As air pollution is a complex mixture of toxic components with considerable impact on humans, forecasting air pollution concentration emerges as a priority for improving life quality. With the rapid development of urbanization, air pollution is becoming a severe environmental and societal issue for all developing countries around the world. Air pollution consists of a mixture of particulate matter and gaseous species (i.e. NO2, CO, O3 and SO2), which have both acute and chronic effects on human health, especially for young and elderly. Machine learning, as one of the most popular techniques, is able to efficiently train a model on big data by using large-scale optimization algorithms. Although there exist some works applying machine learning to air quality prediction, most of the prior studies are restricted to several-year data and simply train standard regression models (linear or nonlinear) to predict the hourly air pollution concentration. Machine learning algorithms become complex because of time and computational complexity. To overcome the drawbacks in existing system, we propose XGBoost deep learning based approach, which consists of a spatial transformation component and a deep distributed fusion network. Considering air pollutants' spatial correlations, the former component converts the spatial sparse air quality data into a consistent input to simulate the pollutant sources to predict the air quality index. We can evaluate the performance of the systems in terms of accuracy for analyzing each and every attributes in Air Datasets.*

**Keywords:** Air Pollution, Particulate Matter(PM2.5)

## I. INTRODUCTION

Pollution of the air is currently one of the most pressing issues facing humanity on a global scale, particularly in highly developed cities. A material from any atmospheric source that exists in any form, whether liquid, solid, or gas phase, and is capable of causing destruction or the capability to alter the typical properties of the atmosphere and increasing the health risk to living things or causing the environment and ecosystem to be out of balance is referred to as polluting the air. PM2.5 is the air pollutant that the United States Environmental Protection Agency (US-EPA) identifies as being the most dangerous to human health and the leading cause of death worldwide. The concentration of air pollutants, particularly PM2.5 during the haze season, is the primary factor that determines the API readings in Malaysia. PM2.5 is a very tiny particle that is able to diffuse into respiratory systems and has a negative impact on human lungs. The weight of PM2.5 is a fine and tiny particle in the air, and its diameter is 2.5 micrometers, as seen in PM2.5 is a very tiny particle that can be visualized in PM2.5. Pollution from vehicles and industrialization are the two primary sources of PM2.5 emissions. Cities that use smart city technology can be observed to be experiencing this stage of the PM2.5 problem. Problems with air pollution are plaguing a number of smart cities.

According to the World Health Organization (WHO), air pollution is a contributing factor in approximately 1.3 million deaths each year around the world. The release of pollutants into the atmosphere has many negative effects, one of which is a deterioration in the quality of the air. Other negative effects, such as acid rain, global warming, the production of aerosols, and photochemical smog, have also worsened over the course of the **past few decades. Many researchers have been motivated to investigate the** underlying pollution-related conditions that are contributing to COVID-19 pandemics in different countries as a result of the recent rapid spread of COVID-19.

Air pollution has been linked to significantly higher COVID-19 death rates, and patterns in COVID-19 death rates mimic patterns in both areas with a high population density and areas with a high PM 2.5 exposure. This is evidenced by several pieces of circumstantial evidence. Because of everything that has been discussed up to this point, it is absolutely necessary to forecast and prepare for changes in pollution levels in order to assist communities and individuals in becoming more effective at mitigating the harmful effects of air pollution. Evaluation of the air's quality is an important factor in both the monitoring and the regulation of pollution levels in the atmosphere.

**OBJECTIVES**

The main objectives of the thesis is to build and train a models using machine learning algorithms and find out the most accurate model in predicting the AQI.

The following are the objectives set to achieve the aim:

- To review studies related to AQI.
- To determine the most usual supervised machine learning algorithm that were used to predict the AQI by performing a literature review.
- To build predictive machine learning models for forecasting AQI.
- To evaluate the performance and accuracy of created models and determine the most accurate supervised machine learning algorithm in the prediction of AQI.

## II. LITERATURE SURVEY

**Nyoman Kusuma Wardana1, Suhaib A. Fahmy and Julian W. Gardner,** designed a Low-cost air quality monitoring devices can provide high-density spatiotemporal pollution data, thus offering a better opportunity to apply machine learning (ML). Low-cost sensor nodes usually utilize microcontrollers as the main processors, and tinyML brings ML models to these resource-constrained devices. In this letter, we report the development of a low-cost air quality monitoring device with embedded tinyML models.

**Yang Han, Jacqueline C.K. Lam , Member, IEEE, Victor O.K. Li , designed a** Predicting air pollution concentration is crucial and beneficial for public health. This study proposes a domain-specificBayesian deep-learning model for long-term air pollution forecast in China and the United Kingdom. Our model outperforms other baseline models. Resultsshow that incorporating Bayesian and domain-specific knowledge into the deep learning model can reduce the prediction errors by amaximum of 3.7% and 12.4%, for Beijing and London,respectively. Specifically, incorporating domain-specific knowledge into theBayesian deep-learning model reduces prediction errors whilst the integration of Bayesian techniques allows the fusion ofdifferentforecast strategies to improve prediction accuracy. In future, additional influential domain-specific features can be added to furtherimprove our deep-learning model's prediction accuracy and interpretability.

**Xiuwen Yi , Zhewen Duan , Ruiyuan Li, Junbo Zhang , proposed a** suffering from air pollution problems, which endangered the health of the young and elderly for breathing problems. For supporting the government's policy-making and people's decision making, it is important to predict future fine- grained air quality. In this article, we predict the air quality of the next 48 hours for each monitoring station and the daily average air quality of the next 7 days for a city, considering air quality data, meteorology data, and weather forecast data. Based on the domain knowledge about air pollution, we propose a deep neural network based approach, entitled DeepAir. Our approach consists of a deep distributed fusion network for station-level short-term prediction and a deep cascaded fusion network for the city-level long-term forecast. With the data transformation preprocessing, the former network adopts a neural distributed architecture to fuse heterogeneous urban data for simultaneously capturing the direct and indirect factors affecting air quality.

**Tianrui Li , Senior Member and Yu ZhengTien Huu Do, Evaggelia Tsiligianni, designed a** Internet-of-Things (IoT) technologies incorporate a large number of different sensing devices and communication technologies to collect a large amount of data for various applications. Smart cities employ IoT infrastructures to build services useful for the administration of the city and the citizens. In this article, we present an IoT pipeline for acquisition, processing, and visualization of air pollution data over the city of Antwerp, Belgium. Our system employs IoT devices mounted on vehicles as well as static reference stations to measure a variety of city parameters, such as humidity, temperature, and air pollution. Mobile measurements cover a larger area compared to static stations; however, there is a trade-off between temporal and spatial resolution. We conduct experiments at different spatial and temporal resolution and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, August 2024

compare with state-of-the-art methods to show the efficiency of our approach. The observed and estimated air pollution values can be accessed by interested users through a Web visualization tool designed to provide an air pollution map of the city of Antwerp.

**Xinghan Xu and Minoru Yoneda desinged a** With the development of the data-driven modeling techniques, using the neural network to simulate the transport process of atmospheric pollutants and constructing PM2.5 time- series prediction model have become a hot topic. The existing data-driven approaches often ignore the dynamical relationships among multiple sites in urban areas, which results in nonideal prediction accuracy. In response to this problem,. And the meteorological information from the monitoring stations is fully utilized, which is beneficial for the performance of the proposed model. Specifically, multilayer LSTM networks can simulate the spatiotemporal characteristics of urban air pollution particles. And using the stacked autoencoder to encode the key evolution pattern of urban meteorological systems could provide important auxiliary information for PM2.5 time-series prediction. In addition, multitask learning could automatically discover the dynamical relationship between multiple key pollution time series and solve the problem of insufficient use of multisite information in the modeling process of the traditional data-driven methods. The simulation results of PM2.5 prediction in Beijing indicate the effectiveness of the proposed method.

## III. METHODOLOGY

Monitoring and preserving air quality has become one of the most essential activities in many industrial and urban areas today. The quality of air is adversely affected due to various forms of pollution caused by transportation, electricity, fuel uses etc. With increasing air pollution, we need to implement efficient air quality monitoring models which collect information about the concentration of air pollutants and provide assessment of air pollution in each area. Hence, air quality evaluation and prediction has become an important research area. The quality of air is affected by multi-dimensional factors including location, time, and uncertain variables. Air quality evaluation is an important way to monitor and control air pollution. The characteristics of air supply affect its suitability for a specific use.In recent years, with the rapid development of India's economy and the continuous improvement of people's quality of life, air pollution caused by a large amount of energy consumption has become increasingly serious. Air quality index (AQI) has become an important basis to measure air quality. At present, the research on air quality assessment and prediction methods has become increasingly active at home and abroad, which is of great significance to guide people's production and life. In this project, we can input the air quality datasets and using the XGBoost model of the deep learning ensemble algorithm is performed on the six pollutant concentrations that currently mainly affect air quality, and the hourly prediction of AQI was achieved.

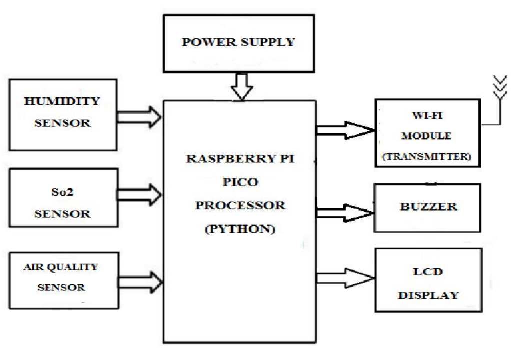


**Fig:1. Block Diagram of the Transmitter**

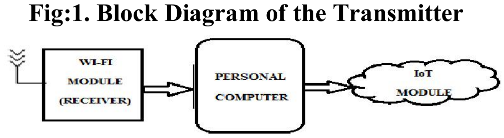


**Fig:2. Block Diagram of the Receiver**

Monitoring and preserving air quality has become one of the most essential activities in many industrial and urban areas today. The quality of air is adversely affected due to various forms of pollution caused by transportation, electricity, fuel uses etc. With increasing air pollution, we need to implement efficient air quality monitoring models which collect information about the concentration of air pollutants and provide assessment of air pollution in each area. Hence, air quality evaluation and prediction has become an important research area. The quality of air is affected by multi-dimensional factors including location, time, and uncertain variables. Air quality evaluation is an important way to monitor and control air pollution. The characteristics of air supply affect its suitability for a specific use.In recent years, with the rapid development of India's economy and the continuous improvement of people's quality of life, air pollution caused by a large amount of energy consumption has become increasingly serious. Air quality index (AQI) has become an important basis to measure air quality. At present, the research on air quality assessment and prediction methods has become increasingly active at home and abroad, which is of great significance to guide people's production and life. In this project, we can input the air quality datasets and using the XGBoost model of the deep learning ensemble algorithm is performed on the six pollutant concentrations that currently mainly affect air quality, and the hourly prediction of AQI was achieved.

**RASPBERRY PI PICO:** A Raspberry Pi Pico is miniature version of the Raspberry Pi that features the RP2040 microcontroller. Raspberry Pi Pico has been developed to be a low-cost and flexible development platform for RP2040
Two megabytes of flash memory on the RP2040 microcontroller;
A micro-USB type B connection for power and data (and for reprogramming the Flash)
40-pin, 2151 "DIP"-style PCB with 0.1" through-hole pins and edge castellations, 1 mm thick
Displays 26 3.3V GPIO that can be used for a variety of purposes.
There are 23 digital GPIO and 3 analog-to-digital converter ones.
Designed to be put on a flat surface like a module

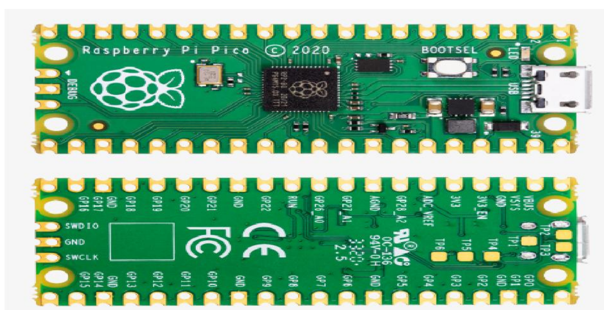


**Figure:1 RASPBERRY PI PICO**

**HUMIDITY SENSOR**: A humidity sensor is an electronic device that measures the humidity in its environment and converts its findings into a corresponding electrical signal. Humidity sensors vary widely in size and functionality; some humidity sensors can be found in handheld devices (such as smartphones), while others are integrated into larger embedded systems (such as air quality monitoring systems). Humidity sensors are commonly used in the meteorology, medical, automobile, HVAC and manufacturing industries.

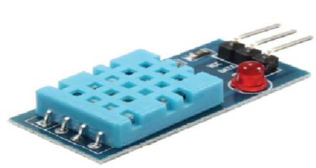

**Figure:2 HUMIDITY SENSOR**

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 1, August 2024

**So2 SENSOR:** A device that is used to detect or measure or monitor the gases like ammonia, benzene, sulfur, carbon dioxide, smoke, and other harmful gases are called as an air quality gas sensor. The MQ135 air quality sensor, which belongs to the series of MQ gas sensors, is widely used to detect harmful gases, and smoke in the fresh air. This article gives a brief description of how to measure and detect gases by using an MQ135 air quality sensor.


**Figure:3 So2 SENSOR or MQ135 Air Quality Sensor**

**LCD DISPLAY:** A **liquid-crystal display** (**LCD**) is a flat-panel display or other electronically modulated optical device that uses the light-modulating properties of liquid crystals combined with polarizers. Liquid crystals do not emit light directly, instead using a backlight or reflector to produce images in color or monochrome. LCDs are available to display arbitrary images (as in a general-purpose computer display) or fixed images with low information content, which can be displayed or hidden, such as preset words, digits, and seven-segment displays, as in a digital clock. They use the same basic technology, except that arbitrary images are made from a matrix of small pixels, while other displays have larger elements. LCDs can either be normally on (positive) or off (negative), depending on the polarizer arrangement. For example, a character positive LCD with a backlight will have black lettering on a background that is the color of the backlight, and a character negative LCD will have a black background with the letters being of the same color as the backlight. Optical filters are added to white on blue LCDs to give them their characteristic appearance.

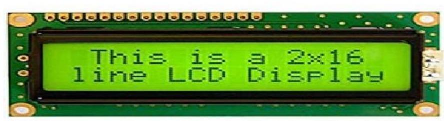


Fig: 4. LCD Display

**POWER SUPPLY:** A power supply (sometimes known as a power supply unit or PSU) is a device or system that supplies electrical or other types of energy to an output load or group of loads. The term is most commonly applied to electrical energy supplies, less often to mechanical ones, and rarely to others

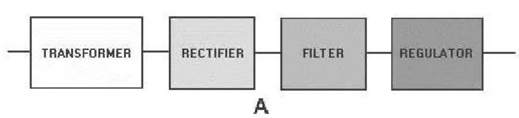


Fig: 5 . POWER SUPPLY

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, August 2024

## V. VERIFICATION AND RESULT

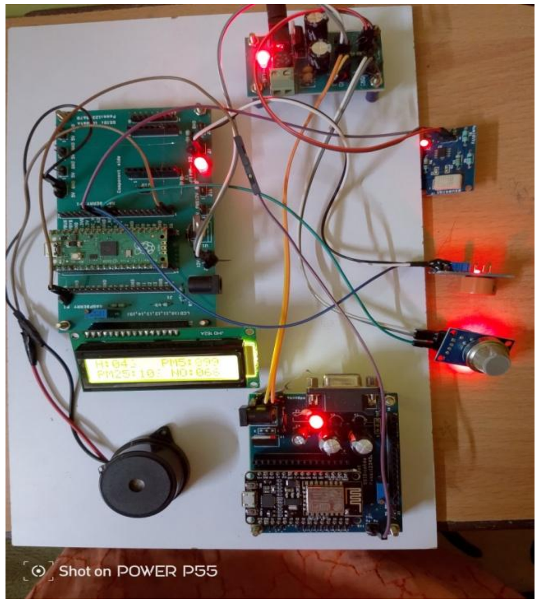


**Fig 1: Prototype of Air quality prediction**


**Fig 2. Simulation Output**

**Fig 3. IoT OUTPUT**

## VI. ADVANTAGES

- Improve the accuracy
- Reduce the computational steps
- Time complexity can be reduced

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, August 2024

## VII. CONCLUSION AND FUTURE SCOPE

Air pollution play hazardous role in the health of the humans and plants. As there are many different sources of pollution, finding the effects of air pollution on health are very complex and their individual effects differ from one to the other. The data are preprocessed and Data can be further processed by data mining tool and proper decision support can be given to the policy makers.

This model also facilitates decision making in day to day life. It can yield better results when applied to cleaner and larger datasets. Pre-processing of the datasets can be effective in the prediction as unprocessed data can also affect the efficiency of the model.

In future this research can be extended to predict the air pollution in various places. This study proves that data mining technology is the optimal solution to design and develop a scalable and secure model to predict the future values and help the policy makers to take the remedial measure to minimize it. Data mining technique when combined with GIS contribute towards the new technology of spatial data mining. Thus survey and analysis of pollution pattern using GIS can become further useful for achieving sustainable management of water resources and air pollutants.

## REFERENCES

**[1].** Soh, Ping-Wei, Jia-Wei Chang, and Jen-Wei Huang. "Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations." Ieee Access 6 (2018): 38186-38199.

**[2].** Fan, Junxiang, et al. "A spatiotemporal prediction framework for air pollution based on deep RNN." ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4 (2017): 15.

**[3].** Kang, Gaganjot Kaur, et al. "Air quality prediction: Big data and machine learning approaches." International Journal of Environmental Science and Development 9.1 (2018): 8-16.

**[4].** Qi, Zhongang, et al. "Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality." IEEE Transactions on Knowledge and Data Engineering 30.12 (2018): 2285-2297.

**[5].** Karimian, Hamed, et al. "Evaluation of different machine learning approaches to forecasting PM2. 5 mass concentrations." Aerosol and Air Quality Research 19.6 (2019):1400-1410.

**[6].** C. J. Wong, M. Z. MatJafri, K. Abdullah, H.S. Lim, and K. L. Low, “Temporal air quality monitoring using surveillance camera,” presented at IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2007

**[7].** Sarkar and P. Pandey, “River water quality modelling usingartificial neural network technique,” Aquatic Procedia, vol. 4, pp. 1070-1077, 2015

**[8].** R. Yu, Y. Yang, L. Yang, G. Han, and, O. A. Move, “RAQ–A random forest approach for predicting air quality in urban sensing systems,” Sensors, vol. 16, no. 1, p. 86, 2016

**[9].** S. Deleawe, J. Kusznir, B. Lamb, and D. J. Cook, “Predicting air quality in smart environments,” J Ambient Intell Smart Environ., pp. 145-152, 2010

**[10].** J. Gao, C.-L. Xie, and C.-Q. Tao, “Big data validation and quality assurance - issues, challenges, and needs,” IEEE Symposium onService-Oriented System and Engineering, IEEE Computer Society, Oxford, UK, April, 2016


DOI: 10.48175/568