Journal of Advanced Zoology
ISSN: 0253-7214
Volume 45 Issue S-4 Year 2024 Page 43-49

Design And Development Of Aerial Vehicle For Air Quality Monitoring

Dr. Janardan Pawar[1*], Madhavi Avhankar[2], Sarika Thakare[3]

[1,2,3] Indira College of Commerce and Science, Pune, Maharashtra, India.*
Email: janardanp@iccs.ac.in[1], madhavi.avhankar@gmail.com[2], sarika.thakare@iccs.ac.in[3]

***Corresponding Author:** Dr. Janardan Pawar*
**Indira College of Commerce and Science, Pune, Maharashtra, India. Email: janardanp@iccs.ac.in*

Abstract

Air pollution is a significant health concern in India. As a developing country, India has numerous issues such as air pollution and other issues. The aim of this study is to design and develop a system that is used to monitor and predict the air quality index (AQI) of the area using machine learning and IoT (Internet of Things). The system will generate AQI values and a line plot graph for future forecasting values. The proposed system will make it simple, convenient, and convenient for users to monitor air quality. It will also be used to forecast the Air Quality Index of the region using the Zhongli F1-AQI. The performance of the proposed system is evaluated using 5 different methods with and without imputation.

Keywords: Air Pollution, Air Quality Monitoring, Internet of Things.

I. INTRODUCTION

As a developing country, India has numerous issues such as air pollution and other environmental issues. Air pollution is a significant health concern in India. According to analytical studies, India had 21 of the top 30 most polluted cities globally. Poor air quality is caused by air pollution, which is a major health concern. Industries, which are the backbone of the country's economic development, are responsible for around 51% of air pollution in the country.

To achieve this, we take the help of machine learning and IoT (Internet of Things). The internet of things (IoT) is a network of networked objects that communicate data over the internet, Wi-Fi, and other wireless technologies. Gas sensors and a Raspberry Pi 4 are the IoT devices that will be used in this example. A flight controller, transmitter, and receiver are the remaining hardware components. The rolling stats and averaging are covered by the machine learning element. Data may be transferred around the globe in seconds thanks to the internet of things.

II. PROBLEM STATEMENT

Some researchers also worked on architectures other than WSN and IoT, but few parameters reveal the low performance of such systems as compared to the potential of IoT systems for real-time monitoring. The most significant disadvantage of the C-Air platform presented by Wu et al.[2] was that this study was limited to PM levels only; but in the real world, IAQ is affected by many other pollutants as well. One of the prime concerns in the development of AQI systems is the accuracy of prediction and the massive power consumption of sensor nodes.

If we consider the real-time applications of AQI systems, the sensor units are usually installed in an industrial environment, inside homes, offices, and outdoor areas as well. However, in all these cases, the design of the sensor unit demands more focus on the accuracy of predicted data, size of the system, processing time, design cost, power consumption, communication protocol, and performance dependence on temperature and humidity variations.

.

III. OBJECTIVES, SCOPE AND IDEA OF THE PROPOSED SYSTEM

OBJECTIVE and SCOPE

- To design and develop a system that is used to monitor and predict the AQI of the area using sensors and predict the AQI for the future time value.
- To evaluate the AQI index of the proposed system.
- To optimize the accuracy, and processing time of the proposed system using Machine learning algorithms and to increase the accuracy of prediction for future AQI
- Comparison of all models using various measures.

IDEA

Overall Algorithm for the proposed System:
Step 1: upload the dataset values to the server
Step 2: collect the values (v) from the target region
Step 3: collected data is uploaded on the server
Step 4: the data (d) is analyzed using moving average algorithm
Step 5: The values (v) are then processed to calculate the AQI based on the formula:

$$I = \frac{I_{high} - I_{low}}{C_{high} - C_{low}} \cdot (C - C_{low}) + I_{low}$$

I: the (Air Quality) index
C: the pollutant concentration
C_{low}: the concentration breakpoint that is <= C
C_{high}: the concentration breakpoint that is >= C
I_{low}: the index breakpoint corresponding to C_{low}
I_{high}: the index breakpoint corresponding to C_{high}

Step 6: set time frame (t) for forecasting
Step 7: Run moving average for the time frame (t) on the data Step 8: send the result of step 7 to the server
Step 9: Visualize step 8 in line plots.

IV. DATA COLLECTION

The main pollutant emissions in India are due to the energy production industry, traffic, waste incineration and agriculture. In India, six pollutants (O3, PM2.5, PM10, CO, SO2, and NO2) are monitored and controlled based on their concentration time series. Types of data used as predictors to perform analysis involve AQ: air quality data, MET: meteorological data, and TIME: the day of the month, day of the week, and the hour of the day. From 1 January 2012 to 31 December 2022, air quality data are collected from several monitoring stations across India and reported via the EPA's website. In the same timeframe, meteorological data are provided in 1-h intervals by India's Central Weather Bureau (CWB) from three air monitoring stations: Zhongli (Northern India), Chuanghua (Central India), and Fengshan (Southern India). The datasets represent different environmental conditions related to air pollutant concentration.

V. DATA PRE-PROCESSING

The number of raw data points for the Zhongli, Changhua, and Fengshan monitoring stations includes 91,672, 94,453, and 94,145, respectively. The analysis of these readings begins with a crucial phase – data preprocessing. Various preprocessing operations precede the learning phase. At any particular time, one invalid variable will not affect the whole data group, and thus it will just be either marked blank or, where available, replaced by a value sourced from the CWB, without eliminating the full row. The missing values are treated by imputation to recover the corresponding values. Given the lack of spatial proximity of the readings to the original monitoring stations, the missing values are imputed for relative humidity, temperature, and rainfall, without using wind speed or wind direction. The next imputation process used the k-NN algorithm to substitute the rest of the invalid or missing data that did not qualify for the previous imputation process. Note that the percentage of missing values is lower than 1.3% in all three-station datasets. Then, input and target data are normalized to eliminate potential biases; thus, variable significance won't be affected by their ranges or their units. All raw data values are normalized to the range of [0, 1] Inputs with a higher scale than others will tend to dominate the measurement and are consequently given greater priority. Normalization not only improves the model learning rate, but also supports k-NN algorithm performance 20 because the imputation is decided by the distance measure. Feature Engineering In regard to selecting features in the predictive models, the hourly AQI readings with the highest index out of 6 pollutants: O3, PM2.5, PM10, CO, SO2, and NO2 are selected. To convert the time-window-specific concentration of 6 pollutants, the AQI India Guidelines [18] are adopted and the AQI is manually calculated using the following Equations (1) and (2), where index values of O3, PM2.5, and PM10 are needed to define AQI in India, and the lack of one or more of these values will significantly reduce the accurate assessment of current air quality.

$$AQI = \begin{cases} Max\{ I_{o3}, I_{PM10}, I_{CO}, I_{SO2}, I_{NO2}\}, I_{o3}, I_{PM2.5}, I_{PM10} \neq Ø \\ Ø \text{ otherwise} \end{cases} \quad (1)$$

Pollutant concentration(value$_i$) is converted to pollutant index(I$_i$) by the following formula:

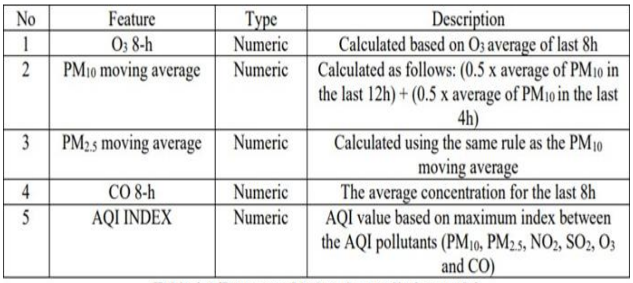
$$I_i = LB_i + \frac{value_i - lb_i}{ub_i - lb_i} \text{ x } (UB_i - LB_i) \quad (2)$$

where i = O3, PM2.5, PM10, CO, SO2, NO2; j denotes which level in the AQI system is occupied by the concentration of the specific pollutant using categories of good, moderate, unhealthy which includes specific groups, unhealthy, very unhealthy, and hazardous. The data transformation defines the time-window-specific concentration to calculate Ii values. For example, based on the AQI from India's EPA website [18], the concentration valueO3 = 0.06 ppm will fall in the interval with lbO3 = 0.055 ppm and ubO3 = 0.070 ppm corresponding to the "moderate" pollutant level with LB moderate = 51 and UB moderate= 100. The valueO3 is defined by matching either of two conditions: if the 8-h average concentration is more precautionary for a specific site and is also below 0.2 ppm, then this value is used; otherwise, the 1-h average concentration will be considered. Both valuePM2.5 and valuePM10 are the moving average values that consider two-time windows, i.e., the last 12 h and 4 h (see Table 1).

Other variables, such as value and valueNO2 only account for a single time window, i.e., last 8 h and 1 h, respectively. Meanwhile, valueSO2 emphasizes the 24-h average concentration if the 1-h average concentration exceeds 185 ppb; otherwise, the 1-h average value will be used. The AQI mechanism introduces several new variables to train the prediction model (Table 1). For several pollutants, time windows other than hourly are more sensitive in determining AQI; therefore, the prediction interval related to the accuracy of long-term predictions is under investigation to clarify the time dependency between consecutive data points. As the AQI calculation is already established, the future value of the AQI readings in three different time intervals will be regarded as target variables and are summarized in Table 2.

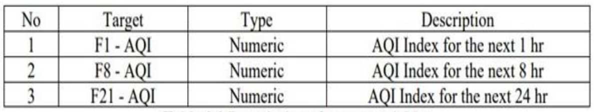
VI. PERFORMANCE EVALUATION

According to Isakndaryan et al., the most used metrics are RMSE (root mean squared error) and MAE (mean average error), calculated based on the difference between the prediction result and the true value, while another metric, R 2 (R-squared) is essential to explain the strength of the relationship between predictive models and target variables [20]. These three metrics provide a baseline for comparative analysis across different parameter settings for each model and across different methods. However, performance validation 21 leads to a bias when the data set is split, trained, and tested only one time. This also means the result drawn from the testing dataset may no longer be valid after the testing subset is changed. To overcome this problem, each model is re-built 20 times using different random subsets of training and testing samples. The splitting proportion remains the same (80:20). All metrics report only a single value from the average performance of 20 identical models validated into 20 different subsets of testing instances.



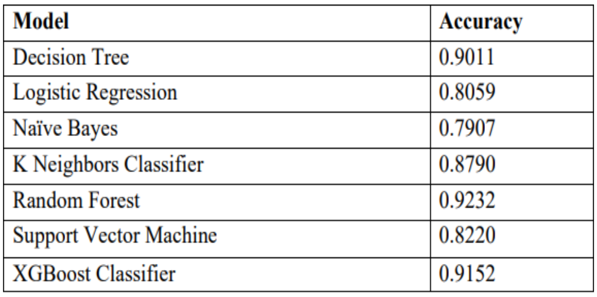
No	Feature	Type	Description
1	O_3 8-h	Numeric	Calculated based on O_3 average of last 8h
2	PM_{10} moving average	Numeric	Calculated as follows: (0.5 x average of PM_{10} in the last 12h) + (0.5 x average of PM_{10} in the last 4h)
3	$PM_{2.5}$ moving average	Numeric	Calculated using the same rule as the PM_{10} moving average
4	CO 8-h	Numeric	The average concentration for the last 8h
5	AQI INDEX	Numeric	AQI value based on maximum index between the AQI pollutants (PM_{10}, $PM_{2.5}$, NO_2, SO_2, O_3 and CO)

Table 3.1 Features added to the prediction model



No	Target	Type	Description
1	F1 - AQI	Numeric	AQI Index for the next 1 hr
2	F8 - AQI	Numeric	AQI Index for the next 8 hr
3	F21 - AQI	Numeric	AQI Index for the next 24 hr

Table 3.2 Description of target variables

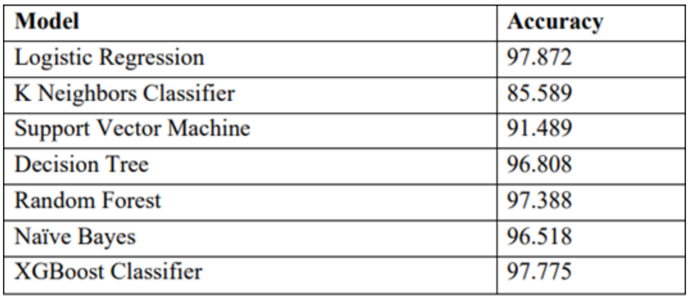


Model	Accuracy
Decision Tree	0.9011
Logistic Regression	0.8059
Naïve Bayes	0.7907
K Neighbors Classifier	0.8790
Random Forest	0.9232
Support Vector Machine	0.8220
XGBoost Classifier	0.9152

Table 3.4 Accuracy on Untrained data

VII. AQI PREDICTION MODEL

The design of the parameters used to generate the prediction models for all datasets. Note that each particular constant for each dataset supposedly contains three values. However, to ease the documentation, any similar value being used across all datasets or at least across different time steps will be written only once. For example, the Changhua dataset uses the number of trees (i.e., 100) in AdaBoost for all time step categories. Additionally, parameter m in the random forest has only one value in all models. To be able to evaluate the ability of each model in accomplishing the task, 80% of data points will be fed into each training process, while the remaining 20% are spared for the testing purpose.

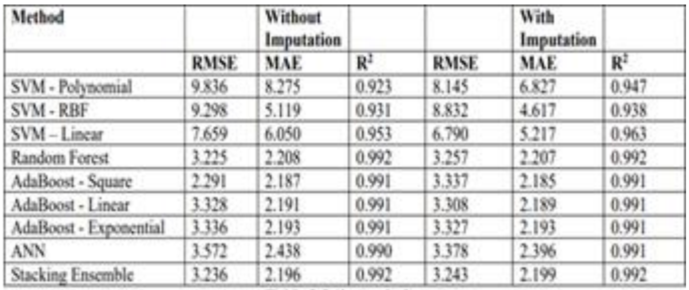


Model	Accuracy
Logistic Regression	97.872
K Neighbors Classifier	85.589
Support Vector Machine	91.489
Decision Tree	96.808
Random Forest	97.388
Naïve Bayes	96.518
XGBoost Classifier	97.775

Table 3.3 Accuracy on Trained dataset

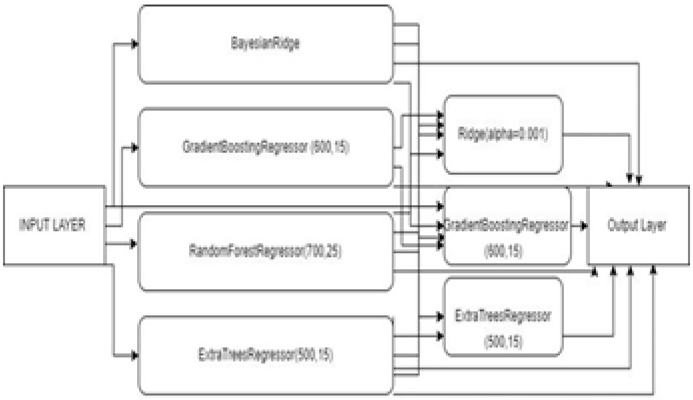
evaluation results of Zhongli F1-AQI prediction using 5 methods with and without imputation. It can be inferred that machine learning algorithms performed very well in predicting future AQI levels in Zhongli for the following hour. The linear kernel is shown to be the best input transformation technique for SVM, with R 2 results of 0.953 (without imputation) and 0.963 (with imputation). Imputation allows SVM to produce improvement in all evaluation metrics.

Furthermore, in terms of MAE score, SVM-RBF outperforms SVMLinear, but the opposite is true for the RMSE score. This may be due to RBF having more samples with a larger prediction error despite a smaller average error (larger errors produce a greater penalty for RMSE).

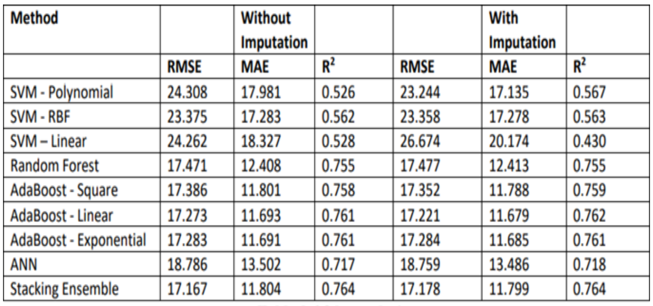


Method		Without Imputation			With Imputation	
	RMSE	MAE	R^2	RMSE	MAE	R^2
SVM - Polynomial	9.836	8.275	0.923	8.145	6.827	0.947
SVM - RBF	9.298	5.119	0.931	8.832	4.617	0.938
SVM – Linear	7.659	6.050	0.953	6.790	5.217	0.963
Random Forest	3.225	2.208	0.992	3.257	2.207	0.992
AdaBoost - Square	2.291	2.187	0.991	3.337	2.185	0.991
AdaBoost - Linear	3.328	2.191	0.991	3.308	2.189	0.991
AdaBoost - Exponential	3.336	2.193	0.991	3.327	2.193	0.991
ANN	3.572	2.438	0.990	3.378	2.396	0.991
Stacking Ensemble	3.236	2.196	0.992	3.243	2.199	0.992

Table 3.5 1hr analysis


BayesianRidge
GradientBoostingRegressor (600,15)
INPUT LAYER
RandomForestRegressor(700,25)
ExtraTreesRegressor(500,15)
Ridge(alpha=0.001)
GradientBoostingRegressor (600,15)
Output Layer
ExtraTreesRegressor (500,15)

The performance of random forest, AdaBoost, ANN, and stacking ensemble algorithm are all comparable. Random forest and stacking ensemble algorithm obtain slightly better R 2 performance (0.001). Unlike with SVM, imputation does not affect the prediction results for AdaBoost, random forest or the stacking ensemble algorithm, indicating their robustness to missing data. On the other hand, imputation only provides a small degree of improvement on ANN, resulting in tied R 2 values with AdaBoost. Several loss regression functions (square, linear, exponential) are tested on AdaBoost but without a decisive performance outcome due to efforts to avoid bias since the interpretation could be distorted by randomness, especially given very minor degrees of difference. Table 5 summarizes the results for the 8-h Zhongli AQI prediction. The R 2 value of 0.764 is the best value obtained by the stacking ensemble method. Nonetheless, the performance of SVM becomes worse with an R 2 value less than 0.6 across all kernels. The values of MAE and RMSE are 17 and 23, respectively. However, ANN and random forest perform better than SVM, with R 2 scores exceeding 0.7 and error metrics just slightly lower than those obtained with AdaBoost and stacking ensemble. The results match the expectation since the uncertainty increases with the longer period and leads to higher difficulty in the forecast. The study also finds that the overall values are worse than that of the F1-AQI prediction.

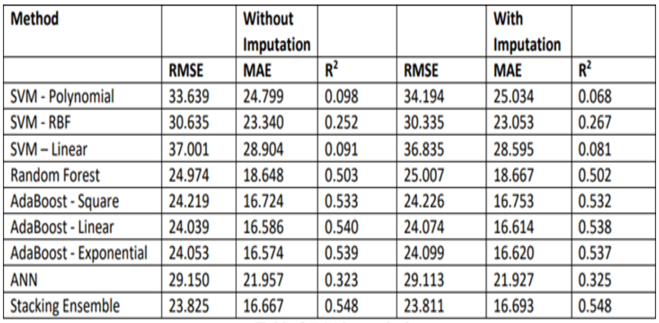


Method		Without Imputation			With Imputation	
	RMSE	MAE	R^2	RMSE	MAE	R^2
SVM - Polynomial	24.308	17.981	0.526	23.244	17.135	0.567
SVM - RBF	23.375	17.283	0.562	23.358	17.278	0.563
SVM – Linear	24.262	18.327	0.528	26.674	20.174	0.430
Random Forest	17.471	12.408	0.755	17.477	12.413	0.755
AdaBoost - Square	17.386	11.801	0.758	17.352	11.788	0.759
AdaBoost - Linear	17.273	11.693	0.761	17.221	11.679	0.762
AdaBoost - Exponential	17.283	11.691	0.761	17.284	11.685	0.761
ANN	18.786	13.502	0.717	18.759	13.486	0.718
Stacking Ensemble	17.167	11.804	0.764	17.178	11.799	0.764

Table 3.6 8 hr analysis

Table 6 shows that no method used for targeting F24-AQI prediction produced an R 2 score above 0.6, with the lowest score of 0.091. Simply put, the yielded predictions fit the dataset poorly. Stacking ensemble still ranks first, but the R 2 gap to the second-best method (AdaBoost-Linear) is larger than in the previous cases. SVM performance is tracked far behind the other methods with the highest score for evaluation metrics obtained by the RBF kernel.

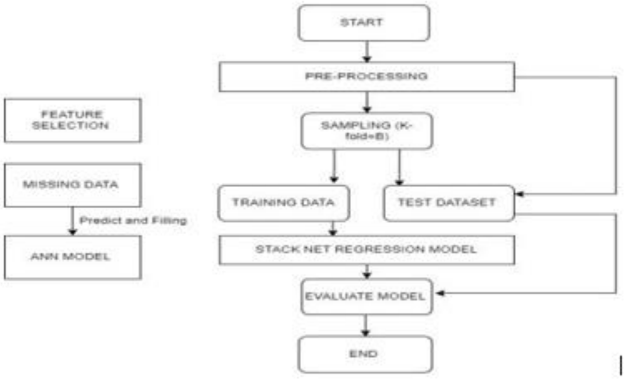
However, the R 2 score is so low that the SVM method is considered not preferable for 24-h prediction.



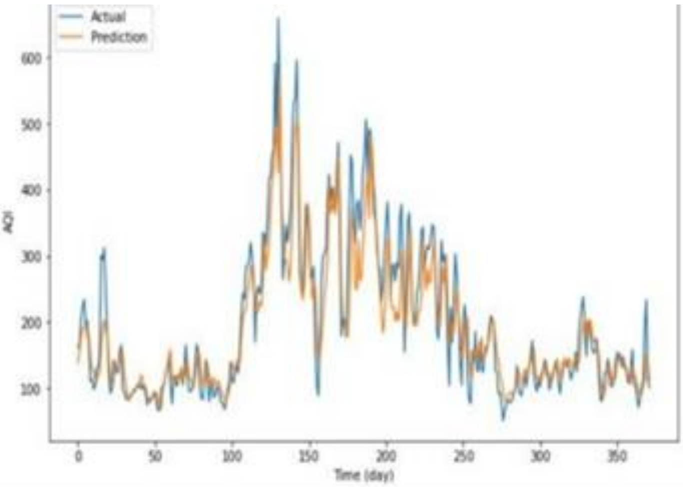
Method		Without Imputation			With Imputation	
	RMSE	MAE	R^2	RMSE	MAE	R^2
SVM - Polynomial	33.639	24.799	0.098	34.194	25.034	0.068
SVM - RBF	30.635	23.340	0.252	30.335	23.053	0.267
SVM – Linear	37.001	28.904	0.091	36.835	28.595	0.081
Random Forest	24.974	18.648	0.503	25.007	18.667	0.502
AdaBoost - Square	24.219	16.724	0.533	24.226	16.753	0.532
AdaBoost - Linear	24.039	16.586	0.540	24.074	16.614	0.538
AdaBoost - Exponential	24.053	16.574	0.539	24.099	16.620	0.537
ANN	29.150	21.957	0.323	29.113	21.927	0.325
Stacking Ensemble	23.825	16.667	0.548	23.811	16.693	0.548

Table 3.7 24 hr analysis

System Architecture



Stacknet Model



.

Results

Applications

1. E-drones offer another way to deal with huge scale air pollutant elimination that will actually automatically monitor and eliminate pollutants that are effectively present in the atmosphere.
2. This approach includes the utilization of drones to independently screen the air quality at a particular area, distinguish the presence of any of these toxins, and carry out an appropriate decrease alternative at a particular elevation to guarantee these pollutants in the atmosphere are removed.
3. The E-drone has been utilized to measure the contamination convergences of O3, CO, NH3, and PM 2.5

VIII. CONCLUSION

Our system will make it simple and convenient for users to monitor and assess air quality. The sensors will collect gas concentration data and send it to the database as soon as the user installs our device in the target zone. The Moving Average method will then be used to assess these values. The system will generate AQI values and a line plot graph for future forecasting values after the analysis is completed. On the user end, these values will be shown on the Thingspeak server. Our approach is inexpensive, effective, and transportable. This lets several users to access and use a single system, with the results accessible from any device with an internet connection.

IX. FUTURE WORK

Many aspects of our work's future development and advancement are still being explored. With the advancement of nanotechnology, the UAV's compactness can be further increased to the point that it can fit into a tiny container. A single mobile application that can be connected to the sensors can replace the need for the internet for analysis. The analysis can now be done more accurately due to advancements in AI and machine learning. The smartphone application can make connecting faraway places easier. Only the overall air index sensor is currently employed to determine patterns and values. With technological advancements, we will be able to use more sensors without compromising the weight capacity. This will result in better analysis using more parameters. India's meteorological department wants to automate the detecting the air first-class is right or no longer from the eligibility method (actual time). To automate this technique by displaying the prediction results in internet software or desktop application. To optimize the paintings to put into effect in an Artificial Intelligence environment.

X. REFERENCES

1. A. D. Deshmukh and U. B. Shinde, “A low cost environment monitoring system using raspberry pi and arduino with zigbee,” in Inventive Computation Technologies (ICICT), International Conference on, vol. 3. IEEE, 2016, pp. 1–6.
2. H. Kumbhar, “Wireless sensor network using xbee on arduino platform: An experimental study,” in Computing Communication Control and automation (ICCUBEA), 2016 International Conference on. IEEE, 2016, pp. 1–5.
3. S. Santini, B. Ostermaier, and A. Vitaletti, “First experiences using wireless sensor networks for air pollution monitoring,” in Proceedings of the workshop on Real-world wireless sensor networks. ACM, 2008, pp. 61–65.
4. M. Mohan and A. Kandya, “An analysis of the annual and seasonal trends of air quality index of delhi,” Environmental monitoring and assessment, vol. 131, no. 1-3, pp. 267–277, 2007.
5. S. Karamchandani, A. Gonsalves, and D. Gupta, “Pervasive monitoring of carbon monoxide and methane using air quality prediction,” in Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on. IEEE, 2016, pp. 2498–2502.
6. U. Z. Jovanovic, I. D. Jovanovic, A. Z. Petrusic, Z. M. Petrusic, and D. D. Mancic,
7. “Low-cost wireless dust monitoring system,” in Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS), 2013 11th International Conference on, vol. 2. IEEE, 2013, pp. 635–638.
8. A. DAusilio, “Raspberry pi: A low-cost multipurpose lab equipment,” Behavior research methods, vol. 44, no. 2, pp. 305–313, 2012. 34
9. R. K. Kodali and A. Sahu, “An IoT based weather information prototype using wemos,” in Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on. IEEE, 2016, pp. 612–616.
10. Avhankar, M. S., Pawar, J., Singh, G., Asokan, A., Kaliappan, S., & Purohit, K. C. (2023, May). Simulation Environment for the I9 Vanet Platform. In 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-8). IEEE.
11. Lambey, V., & Prasad, A. D. (2021). A review on air quality measurement using an unmanned aerial vehicle. Water, Air, & Soil Pollution, 232, 1-32.
12. Avhankar, M. S., Pawar, J., & Byagar, S. (2022, December). Localization Algorithms in Wireless Sensor Networks: Classification, Case Studies and Evaluation Frameworks. In 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) (pp. 01-07). IEEE.
13. Camarillo-Escobedo, R., Flores, J. L., Marin-Montoya, P., García-Torales, G., & Camarillo-Escobedo, J. M. (2022). Smart multi-sensor system for remote air quality monitoring using unmanned aerial vehicle and LoRaWAN. Sensors, 22(5), 1706.