Grenze International Journal of Engineering and Technology, June Issue

Predicting Air Quality with Mobile Camera using ML

Pravin Kumbhar[1], Vinay Dasmane[2], Sarang Mashale[3], Vaishnavi Dudhanikar[4] and Prathamesh Kumbhar[5]
[1-5] E&TC Department, Walchand Institute of Technology, Solapur, India
Email: pykumbhar@witsolapur.org, {vinaydasmane, sarangmashale21, dudhanikarvaishnavi, prathameshskumbhar622001}@gmail.com

***Abstract*—Increasing rate of air pollution, have an obvious consequences causing diseases and increasing rate of death of population. The common air pollutants such as CO, SO_2, lead, NO_2, O_3 and PM are the criteria pollutants which corresponds to the most widespread health threats. In this paper, Image processing and Machine learning modules are introduced to predict Air Quality Index (AQI) using camera images. Simulations and visualizations are carried out in Anaconda/Jupiter Notebook. Results provide approximate values of AQI of the given input image.**

***Index Terms*— Air Quality Index (AQI), Image Processing, Machine Learning (ML).**

I. INTRODUCTION

Along with technological and economical growth of cities there is a rise in environmental pollutions such as water, air and noise pollution. Particularly the air pollutions has a wide spread impact on human health due to exposure to primary pollutants such as CO (carbon monoxide), SO_2 (sulfur dioxide), lead, NO_2 (nitrogen dioxide), O_3 (Ozone) and PM (particulate matters). Thus, there is a need to spread awareness among the people regarding the bad quality of air they breathe.

Globally, the use of various AQI include synergistic effect estimation based on mean of the ratio of pollutant over guideline levels for a certain amount of time period. AQI is simplest and widely used measurable tool of overall air pollution of a region. The pollutant with highest proportion in air decides overall AQI for that hour/day.

The Central Pollution Control Board (CPCB) has defined the national Air Quality Index (AQI) for the different cities in India. AQI is calculated using 12 parameters or pollutants namely NO2 (nitrogen dioxide), SO2 (sulphur dioxide), CO (carbon monoxide), O3 (ozone), PM10 (particulate matter having diameter less than or equal to 10 micron), PM2.5 (particulate matter having diameter less than or equal to 2.5 micron), NH3 (ammonia), Pb (lead), Ni (nickel), As (arsenic), Benzo(a)pyrene and Benzene. Here the selection of pollutants depends on the AQI objectives, data availability, averaging µ period, monitoring frequency and measurement methods. Thus AQI can be defined as it is a numerical value that the governmental agencies used to measure the levels of air pollution in the atmosphere and communicate it with population.[1]

The main motive of our idea is to build an android application which can predict AQI to help people know about the poor air quality which can cause some serious health issues. Also to calculate AQI through images captured from user's device using Image processing and Machine learning algorithms.

II. PREVIOUS AND RELATED WORKS

Anikendar Kumar, Pramila Goyal (2011) presented the study which forecast the daily AQI value of the city

Grenze ID: 01.GIJET.8.2.508_5

Delhi, India using previous records of AQI and meteorological parameters using the Principle Component Regression (PCR) and Multiple Linear Regression Techniques. AQI prediction of the year 2006 on the daily basis was performed by them using previous records of the year 2000-2005.Using Multiple Linear Regression Technique, the predicted value is compared with observed value of AQI of 2006 for the seasons summer, monsoon, post monsoon and winter. To eliminate collinearity among predicted variables, principle components were used in Multiple Linear Regression. In this study only meteorological parameters were considered or used while forecasting the future AQI but they have not considered the ambient air pollutants that may cause the adverse health effects.

Aditya C R (et al.2018) has employed the machine algorithms to forecast and detect the PM2.5 concentration level on the basis of dataset having atmospheric conditions in a specific city. They have also predicted the concentration level of PM2.5 of a specific date. Firstly they have classified that the air is polluted or not polluted by using Logistic Regression Algorithm and then Auto Regression Algorithm is used to predict the future value of PM2.5 depending upon previous records.

Nidhi Sharma (et al.2018) has gone through detailed study of data analysis of the air pollutants from the year 2009-2017 and has also proposed the critical observations of air pollutant trends in year 2016-2017 in Delhi, India. Further they have predicted the future trends of various air pollutants. By using data analytics Time Series Regression forecasting they have predicted the future values of pollutants on the basis of previous records. Two monitoring stations in Delhi namely Anand Vihar and Shadipur were under their study. The results have shown that there is drastic increase in concentration level of PM10, NO2 and PM2.5 which has resulted in rise of pollution in Delhi.

Mohamed Shakir and N.Rakesh (2018) have both analyzed the proportion of various air pollutants with respect to the time of the day and the day of the week and observed the effect of environmental parameters like windspeed, temperature and humidity on the air pollutants with the help of WEKA tools. The data was used from the Control Board of Pollution, Karnataka. Using the ZeroR algorithm in WEKA tool, the study shows that the concentration levels of air pollutants have increased during the work days and especially at peak during the hours of the day and slowly decreases during weekends or holidays. With the help of Simple K-means Clustering Algorithms the study resulted the relationship or dependencies between the environmental factors such as humidity, windspeed and temperature and various air pollutants.

III. PROPOSED METHODOLOGY

To understand the temporal variation and episodic rise of the air pollution in the surrounding, real time values of various air pollutants is monitored and used to calculate Air Quality Index. In proposed study, two different modules were combinedly used to determine the actual value of AQI.

MODULE 1: IMAGE PROCESSING MODULE

In this module, the image provided by the user is given to the module as input. Firstly, the model will predict percentage of skyline in the image, if the skyline is more than 50%, image will be forwarded to further model. Second model will extract blue colour of the sky. Due to the Light scattering concept, sky is blue on a clear, non-cloudy day, whereas it appears grey or dark in shade on a hazy and cloudy day. Average value of blue components of RGB channels in sky can be used to find the blue colour of the sky. This will also help in classifying whether image is day or night image.[2].

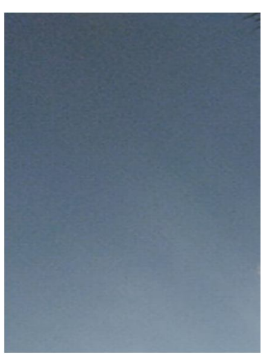

Figure 1. Input Image to the Image based model   Figure 2. Blue colour detection model

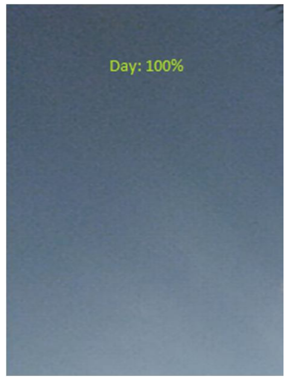

Figure 3. Day/Night Classifier model

MODULE2: PARAMETER BASED CALCULATION OF AQI

This module will fetch values of parameters like PM2.5, PM10, NOx, SO2, CO and Temperature from the dataset provided by cpcb.nic.in site of user's nearest monitoring station. Average values in last 24 hours of PM2.5, PM10, NOx, SO2 and maximum values of CO and Temperature in last 8 hours are passed as input to a trained module to calculate their respective sub-index values.[3] Further, these values are collectively passed to a Random Forest Classifier which will predict current value of AQI.[4-5].

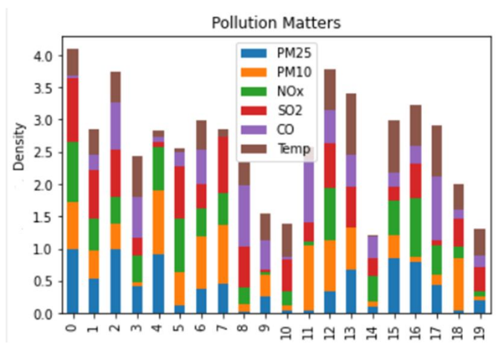

Figure 4. Distribution of all pollutants and AQI

Finally these two modules are combined using Adaptive boosting algorithm to predict current AQI value. Once the dashboard is deployed over a cloud and all the collected data is processed by a trained model AQI value is displayed on screen of the application. Also, the user will be able to view all the previous AQI results and compare them with current AQI.

IV. Results

For the calculation of AQI, large amount of dataset was collected (as shown below) to train machine learning module.

	S.No	PM25	PM10	NOx	SO2	CO	Temp	AQI
count	2263.000000	1994.000000	2028.000000	1766.000000	1970.000000	2081.000000	1933.000000	1078.000000
mean	1132.000000	34.755271	94.385720	56.051444	13.472543	1.388385	28.530595	91.615955
std	653.416151	21.002097	50.773501	39.080953	5.273227	1.620476	5.097084	41.475558
min	1.000000	0.920000	1.120000	0.000000	0.400000	0.000000	0.190000	23.000000
25%	566.500000	15.362500	49.640000	32.082500	10.520000	0.450000	26.560000	53.000000
50%	1132.000000	32.810000	90.160000	45.850000	12.580000	1.250000	28.330000	93.000000
75%	1697.500000	49.667500	131.892500	72.220000	16.235000	2.060000	30.750000	122.000000
max	2263.000000	157.280000	362.520000	402.780000	33.350000	48.460000	41.670000	247.000000

Figure 5. Sample Dataset for AQI calculation

Data collected from the source has more null values since few of the data from the source station have been shifted or closed over a period and those values were marked as NaN or not available.

```
AQI           52.364118
NOx           21.961997
Temp          14.582413
SO2           12.947415
PM25          11.886876
PM10          10.384445
CO             8.042422
City           0.000000
To Date        0.000000
From Date      0.000000
S.No           0.000000
```

Figure 6. Percentage of noisy data in Dataset

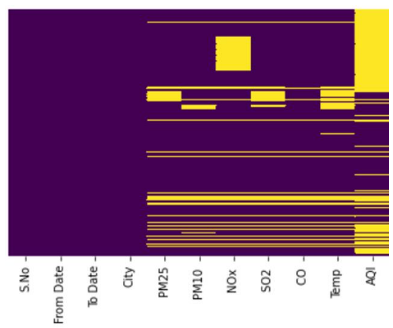

Figure 7. Heatmap of dataset with null values

These NaN values are filled by using interpolate method which calculate the values by taking mean of previous and next value.

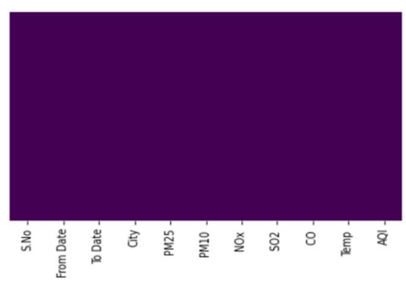

Figure 8. Heatmap of Dataset after filling null values

Distribution of each parameters over a period of time is given below:

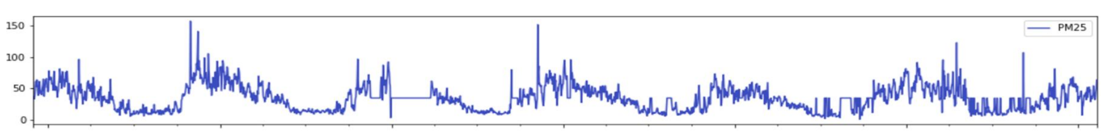

Figure 9. Distribution of PM2.5

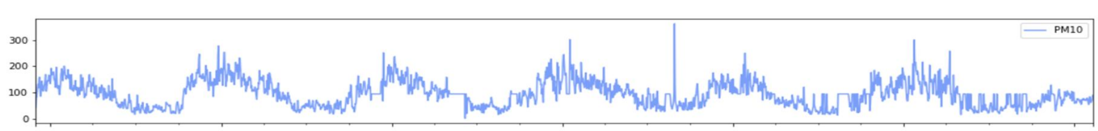

Figure 10. Distribution of PM10

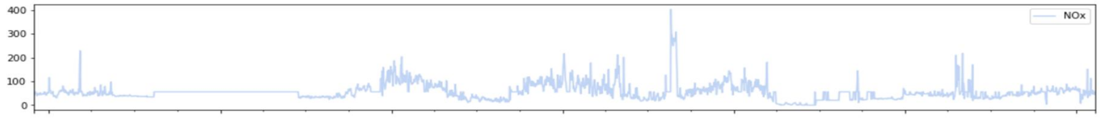

Figure 11. Distribution of NOx

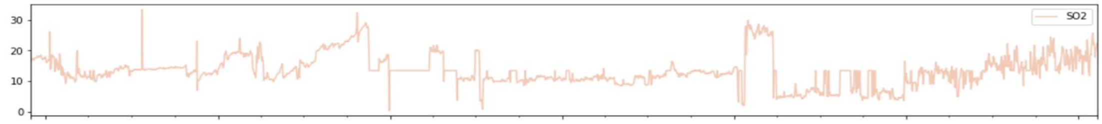

Figure 12. Distribution of SO2

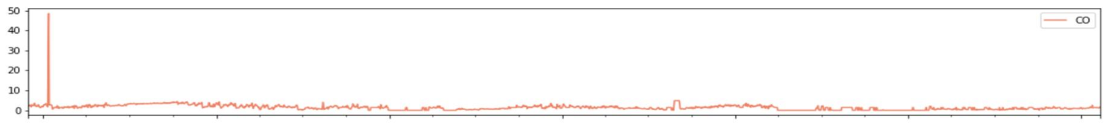

Figure 13. Distribution of CO

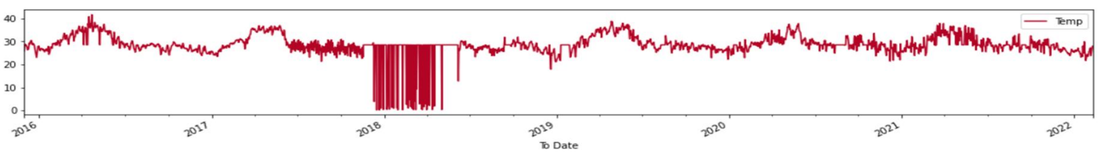

Figure 14. Distribution of Temperature

Subindex of each Parameter is required to calculate the value of AQI. Subindex of each parameter is calculated individually using below code.

```python
# Parameters Sub-Index calculation
def get_parameter_subindex(x):
    if x <= 30:
        return x * 50 / 30
    elif x > 30 and x <= 60:
        return 50 + (x - 30) * 50 / 30
    elif x > 60 and x <= 90:
        return 100 + (x - 60) * 100 / 30
    elif x > 90 and x <= 120:
        return 200 + (x - 90) * 100 / 30
    elif x > 120 and x <= 250:
        return 300 + (x - 120) * 100 / 130
    elif x > 250:
        return 400 + (x - 250) * 100 / 130
    else:
        return 0

df["parameter_SubIndex"] = df["parameter"].astype(int).apply(lambda x: get_parameter_subindex(x))
```

Figure 15. Formula for Subindex calculation

Subindex columns are then added to the dataset.

	S.No	From Date	To Date	City	PM25	PM10	NOx	SO2	CO	Temp	AQI	PM2.5_SubIndex	PM10_SubIndex	SO2_SubIndex	NOx_SubIndex	CO_SubIndex	Temp_SubIndex
0	1	30-Nov-2015 - 00:00	2015-12-01	solapur	34.755271	43.15	42.11	16.33	2.21	25.88	91.615955	56.666667	43.000000	20.00	52.50	100.0	347.058824
1	2	01-Dec-2015 - 00:00	2015-12-02	solapur	37.100000	79.41	45.38	17.20	2.11	28.50	91.615955	61.666667	79.000000	21.25	56.25	100.0	364.705882
2	3	02-Dec-2015 - 00:00	2015-12-03	solapur	33.380000	81.92	54.86	17.46	2.39	28.77	91.615955	55.000000	81.000000	21.25	67.50	100.0	364.705882
3	4	03-Dec-2015 - 00:00	2015-12-04	solapur	39.490000	102.97	56.69	17.30	2.33	28.50	91.615955	65.000000	101.333333	21.25	70.00	100.0	364.705882
4	5	04-Dec-2015 - 00:00	2015-12-05	solapur	54.280000	121.46	47.90	17.08	2.70	28.53	91.615955	90.000000	114.000000	21.25	58.75	100.0	364.705882
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2258	2259	04-Feb-2022 - 00:00	2022-02-05	solapur	32.160000	61.57	61.92	17.94	1.42	23.78	74.000000	53.333333	61.000000	21.25	76.25	50.0	335.294118
2259	2260	05-Feb-2022 - 00:00	2022-02-06	solapur	33.330000	64.11	47.84	19.63	1.23	25.03	65.000000	55.000000	64.000000	23.75	58.75	50.0	347.058824
2260	2261	06-Feb-2022 - 00:00	2022-02-07	solapur	44.720000	68.40	54.61	20.04	1.01	26.04	65.000000	73.333333	68.000000	25.00	67.50	50.0	352.941176
2261	2262	07-Feb-2022 - 00:00	2022-02-08	solapur	49.350000	80.94	44.60	21.74	1.30	27.64	96.000000	81.666667	80.000000	26.25	55.00	50.0	358.823529
2262	2263	08-Feb-2022 - 00:00	2022-02-09	solapur	63.540000	89.42	55.82	22.33	1.62	27.98	88.000000	110.000000	89.000000	27.50	68.75	50.0	358.823529

2263 rows × 17 columns

Figure 16. Dataset including Subindex of each parameter

Below chart gives the different range of AQI and its respective effects on the health of human population

Good (0–50)	Minimal Impact	Poor (201–300)	Breathing discomfort to people on prolonged exposure
Satisfactory (51–100)	Minor breathing discomfort to sensitive people	Very Poor (301–400)	Respiratory illness to the people on prolonged exposure
Moderate (101–200)	Breathing discomfort to the people with lung, heart disease, children and older adults	Severe (>401)	Respiratory effects even on healthy people

Figure 17. AQI Bucket Chart

Some AQI Bucket calculated with respect to actual AQI are given below:

	AQI	AQI_calculated	AQI_bucket_calculated
0	91.615955	347.0	Very Poor
1	91.615955	365.0	Very Poor
2	91.615955	365.0	Very Poor
3	91.615955	365.0	Very Poor
4	91.615955	365.0	Very Poor
...	...	...	...
2258	74.000000	335.0	Very Poor
2259	65.000000	347.0	Very Poor
2260	65.000000	353.0	Very Poor
2261	96.000000	359.0	Very Poor
2262	88.000000	359.0	Very Poor

Figure 18. AQI Bucket Calculated

Heatmap representation of AQI calculated based on the parameters in the dataset :

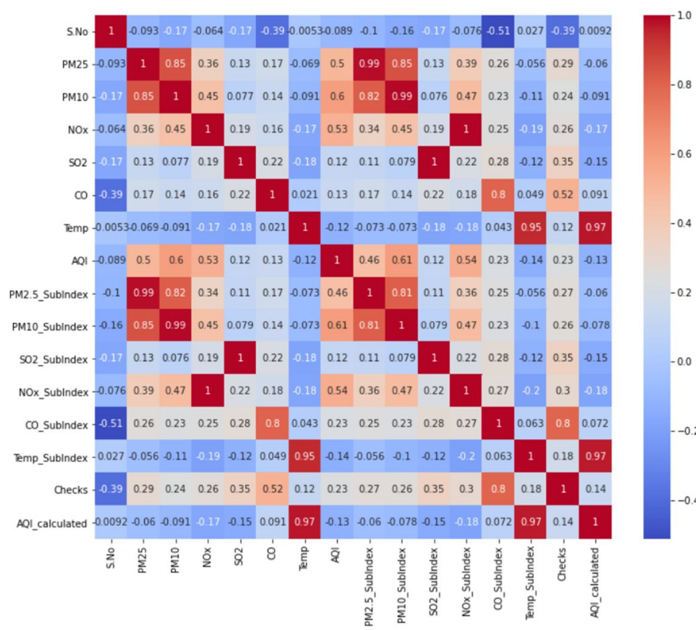

Figure 19. Heatmap of calculated AQI

V. Conclusion

Air quality index gives a clear scenario about the atmosphere and the major pollutants responsible for increasing bad quality of Air. AQI was predicted as per concentration provided by CPCB. This study shows that the pollutant with major proportion in the air decides the value of Air Quality Index. Also particulate matter causes a serious effect on health of population. We have to continuously monitor the quality of air and take appropriate measures to make air breatheable for everyone. This novel approach aims to build an app, which can predict AQI through the images captured from users' mobile and using their location. This idea focuses on visualizing AQI value using Random Forest Classifier Algorithm. This AQI can be visualized on the app which is to be built and also the user will be able to do graphical analysis. For this, two modules namely image based module and parameter based modules are used collectively to predict AQI. Also app development is done using python and kivymd modules. This application will help people to be aware about the air they breathe and help them take precautions and reduce outdoor activities.[6].

REFERENCES

[1] B. R. Gurjar, L. T. Molina, and C. S. P. Ojha, "Air pollution: health and environmental impacts." *CRC press,* 2010.

[2] Park, H.-J.; Jang, J.-I.; Kim, B.-G., "Time Classification Algorithm Based on Windowed-Colour Histogram Matching." Applied Sciences 2021,11, 11997.

[3] Kumar A. , Goyal P. et al, "Forecasting of Air Quality in Delhi using principle component technique." *Atmospheric Pollution Research* 2.2011:436-444.

[4] Bhanarkar A. D., et al, "Assessment of contribution of SO2 and NO2 from different sources in Jamshedpur region, India, *Atmospheric Environment* 39.40(2005): 7745-India.

[5] A. G. Soundari, J. G. Jeslin, Akshaya A. C., "Indian Air Quality Prediction And Analysis Using Machine Learning", *International Journal of Applied Engineering Research ISSN* 0973-4562 Volume 14, Number 11, 2019.

[6] F. Caiazzo, A. Ashok, I. A. Waitz, S. H. L. Yim, and S. R. H. Barrett, "Air pollution and early deaths in the United States. Part I: quantifying the impact of major sectors in 2005", *Atmospheric Environment,* vol. 79, pp. 198–208, 2013.