

Title : A real-time dataset of air quality index monitoring using IoT and machine learning in the perspective of Bangladesh

Author Names : Md. Monirul Islam, Ferdaus Anam Jibon, M. Masud Tarek, Muntasir Hasan Kanchan, Shalah Uddin Perbhez Shakil

Abstract: This paper produces a real-time air quality index dataset of three places named Kuril Bishow Road, Uttara, and Tongi in Dhaka and Gazipur City, Bangladesh. The IoT framework consists of MQ9, MQ135, MQ131, and dust or PM sensors with an Arduino microcontroller to collect real data on sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone, particle matters 2.5 and 10 μm . The data is stored in an Excel file as a comma-separated file and after that, authors applied regression type and classification type machine learning algorithms to analyze the data. The dataset consists of 11 columns and 155,406 rows, where sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone, and particle matter 2.5 and 10 are recorded where AQI is marked as the target variable and the others are indicated as independent variables. In the dataset, AQI is categorized into five classes named Good, satisfactory, Moderate, Poor and Very Poor. After experimental results, it is seen that two places including Uttara and Kuril are comparatively suitable for Air Quality among the three places as well as the Random Forest algorithm outperforms the models. The study describes details of the embedded system's hardware as well. This dataset will be beneficial for environmental researchers to use to analyze the air quality.

Keywords: Air quality index (AQI), Air pollution, Machine learning, IoT sensors

[DOI: 10.17632/4r25x9sc7k.1](https://doi.org/10.17632/4r25x9sc7k.1)

<https://www.sciencedirect.com/science/article/pii/S2352340924005456>