# # Air-Pollution-Image-Dataset-From-India-and-Nepal

# Introduction:

This dataset contains images of Air Pollution for different cities in India and Nepal.

The dataset is divided into two folders: Combined\_Dataset and Country\_wise\_Dataset.

Total number of image dataset: 12,240

## Air Quality Index (AQI) Class

There is a total of six classes of Air Pollution, which we represent in our dataset as follows:

1. a\_Good

2. b\_Moderate

3. c\_Unhealthy\_for\_Sensitive\_Groups

4. d\_Unhealthy

5. e\_Very\_Unhealthy

6. f\_Severe

### Cities of India

1. ITO, Delhi

2. Dimapur, Nagaland

3. Spice Garden, Bengaluru

4. Knowledge Park III, Greater Noida

5. New Ind Town, Faridabad

6. Borivali East, Mumbai

7. Oragadam, Tamil Nadu

### City of Nepal

1. Biratnagar

### Combined dataset:

The combined dataset folder contains two subfolders.

1. All\_img: This subfolder contains all the collected images from all AQI classes.

2. IND\_and\_NEP: This subfolder contains six different subfolders representing six different classes of AQI.

The csv file in this folder contains all the data and its parameters.

It is labeled as

Location, Filename, Year, Month, Day, Hour, AQI, PM2.5, PM10, O3, CO, SO2, NO2, and AQI\_Class

### Country\_wise\_Dataset:

This folder contains two subfolders representing the countries from which the dataset was collected.

1. India:

This subfolder contains the subfolder representing the names of all cities from where data were collected.

Each subfolder of cities contains folders representing the data collected for each respective AQI class, as well as a csv file.

which contains the details of each image, like we mentioned above.

Such as,

Location, Filename, Year, Month, Day, Hour, AQI, PM2.5, PM10, O3, CO, SO2, NO2, and AQI\_Class

2. Nepal:

We managed to collect the image dataset from Nepal.

This subfolder contains the subfolder representing the name of the city from where data were collected.

This subfolder of the city contains folders representing the data collected for each AQI class and also a csv file.

which contains the details of each image, like we mentioned above.

Such as,

Location, Filename, Year, Month, Day, Hour, AQI, PM2.5, PM10, O3, CO, SO2, NO2, and AQI\_Class

## Instructions on how to use the AQI image dataset:

1. Download the dataset from Kaggle and extract the zip file to a folder of your choice. Please Visit this link to download the Dataset:

2. The dataset is divided into two folders: The Combined\_Dataset and Country\_wise\_Dataset.

Each folder contains subfolders and CSV files.

3. To access the images in the Combined\_Dataset folder, go to the folder corresponding to the class of AQI you are interested in.

For example, if you are interested in the 'Unhealthy' class, go to the 'Unhealthy' folder. Inside this folder,

You will find a number of images representing different cities.

4. To access the data in the Country\_wise\_Dataset folder, go to the folder of the country you are interested in, either India or Nepal.

Inside each country folder, you will find subfolders representing different cities.

Each city folder contains a CSV file that lists the AQI values and other parameters for the city.

5. You can use this dataset to train machine learning models to predict AQI for different cities.

You can also use it for research on air pollution in different cities.

6. If you use this dataset for any purpose, please cite it as the source of the data in any publications or presentations, resulting from the use of this dataset.

***7. GitHub Repository:*** [***https://github.com/ICCC-Platform/Air-Pollution-Image-Dataset-From-India-and-Nepal/edit/main/README.md***](https://github.com/ICCC-Platform/Air-Pollution-Image-Dataset-From-India-and-Nepal/edit/main/README.md)

***If you are visiting our GitHub Repository please fork our repository if you find it useful.***

***Forking our repository allows you to create your own copy of our repository, which you can modify and use as you wish.***

***\*\*\*8. Also star our repository if you like it.***

***Starring our repository is a way for people to show their support and appreciation for our work.***

**Citation Request: You can cite our dataset as follows**

APA:

Adarsh Rouniyar, Sapdo Utomo, John A, & Pao-Ann Hsiung. (2023). *Air Pollution Image Dataset from India and Nepal* [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DS/3152196>

Bibtex:

@misc{adarsh rouniyar\_sapdo utomo\_john a\_pao-ann hsiung\_2023, title={Air Pollution Image Dataset from India and Nepal}, url={[https://www.kaggle.com/ds/3152196}](https://www.kaggle.com/ds/3152196%7D), DOI={10.34740/KAGGLE/DS/3152196}, publisher={Kaggle}, author={Adarsh Rouniyar and Sapdo Utomo and John A and Pao-Ann Hsiung}, year={2023} }

## Contributors

1. Adarsh Rouniyar

2. Sapdo Utomo

3. Dr. John A.

4. Dr. Pao-Ann Hsiung

## If you have any queries, please do contact us.

1. Adarsh Rouniyar

Email: [adarsh@csie.io](mailto:adarsh@csie.io)

2. Dr. John A.

Email: [johnmtech@gmail.com](mailto:johnmtech@gmail.com)