

Computer Vision Driven Air Pollution Monitoring using Visual Atmospheric Analysis

Lakshman Kumar, Payal Jha, Laksay Singh Bisht

Department of Computer Science and Engineering, IILM University, Greater Noida, India

Abstract—Air pollution has emerged as one of the most critical global threats to public health and the environment. Conventional monitoring approaches rely on costly, sparsely deployed sensor stations, leaving large geographic regions without reliable air quality information. This paper proposes a deep-learning based framework that estimates air quality directly from atmospheric images using a Convolutional Neural Network (CNN). Visual cues such as haze density, smoke, and reduced visibility are extracted automatically and used to classify ambient air quality into Good, Moderate and Hazardous categories. The model is implemented in TensorFlow/Keras and deployed through a lightweight Flask web application that provides real-time predictions from user-uploaded images. Experimental evaluation on a labelled dataset and on real-world photographs captured in Rohini, New Delhi, demonstrates a classification accuracy of 85–88%, with training and validation accuracy converging above 0.97 within twenty epochs. The proposed system offers a scalable, low-cost complement to hardware-based monitoring and is well suited for smart-city deployment.

Index Terms—Air pollution detection, convolutional neural network, deep learning, TensorFlow, Flask, smart city, AQI estimation, computer vision.

I. INTRODUCTION

Air pollution has become one of the most pressing global concerns of the present decade. Rapid industrialisation, unplanned urban expansion and the continual rise in vehicular traffic have dramatically increased the concentration of harmful pollutants such as PM_{2.5}, PM₁₀, carbon monoxide (CO), sulphur dioxide (SO₂) and nitrogen oxides (NO_x) in ambient air. Prolonged exposure to these substances is strongly correlated with respiratory and cardiovascular diseases and, in severe cases, premature mortality.

Beyond human health, polluted air degrades ecosystems, reduces atmospheric visibility, damages agricultural yield and accelerates climate change. Despite the severity of the problem, comprehensive monitoring of air quality across all regions remains a major challenge.

Traditional air-quality monitoring relies on hardware-based sensor stations installed at fixed sites. While accurate, these stations are expensive to deploy and maintain, and their limited density results in poor spatial coverage—particularly in developing regions and densely populated urban districts.

Recent advances in artificial intelligence, and in particular Convolutional Neural Networks (CNNs), have opened new opportunities for environmental sensing. CNNs automatically extract hierarchical visual features such as edges, textures and

colour gradients without manual feature engineering, making them well suited to analysing atmospheric imagery.

In this work, instead of depending exclusively on physical sensors, images of the sky and surrounding scene are used as the primary input modality. Visual indicators such as haze, smoke and reduced visibility carry rich information about pollution levels. By learning these patterns from labelled data, the proposed system classifies air quality into Good, Moderate or Hazardous categories and is delivered through a lightweight Flask web interface.

A. Motivation

The rapid increase of pollution in urban and developing regions creates an urgent need for affordable and accessible monitoring. Conventional monitoring equipment is costly and not widely deployed, leaving most citizens unable to assess the air they breathe. Modern deep-learning frameworks make it possible to build intelligent vision-based systems that are simple, accurate, low-cost and easy to use. Using TensorFlow and Flask, real-time predictions can be generated and disseminated to the public, raising awareness and supporting better environmental decisions.

B. Our Contributions

- We design a lightweight CNN architecture that classifies ambient air quality into Good, Moderate and Hazardous categories using only visual input.
- We curate and pre-process a labelled image dataset that captures haze, smoke and visibility patterns under varied lighting and weather conditions.
- We deploy the trained model through a Flask-based web service, enabling real-time image-based AQI estimation without dedicated sensing hardware.
- We validate the system on real-world photographs taken in Rohini, New Delhi, with concurrent ground-truth AQI readings, achieving 85–88% classification accuracy.

II. RELATED WORK

Conventional air-quality monitoring depends on physical sensors that directly measure pollutant concentrations such as CO, SO₂, NO_x and particulate matter. While such systems deliver accurate point measurements, they suffer from high installation and maintenance costs and limited spatial density, restricting continuous monitoring across large or remote regions.

To overcome these limitations, several studies have explored machine-learning and deep-learning techniques for air-quality estimation. Earlier methods used historical pollutant time-series with regression and recurrent models, while more recent work focuses on image-based analysis with Convolutional Neural

Networks. Such vision systems detect haze, smoke and visibility degradation from environmental imagery and deliver scalable, accessible alternatives to hardware sensing while maintaining competitive accuracy.

III. PROPOSED METHODOLOGY

A. System Overview

The proposed pipeline, summarised in Fig. 1, consists of four major stages: (i) image acquisition and pre-processing, (ii) feature extraction through stacked convolution and pooling layers, (iii) classification via fully connected layers, and (iv) deployment through a Flask web interface that returns the predicted air-quality class to the user.

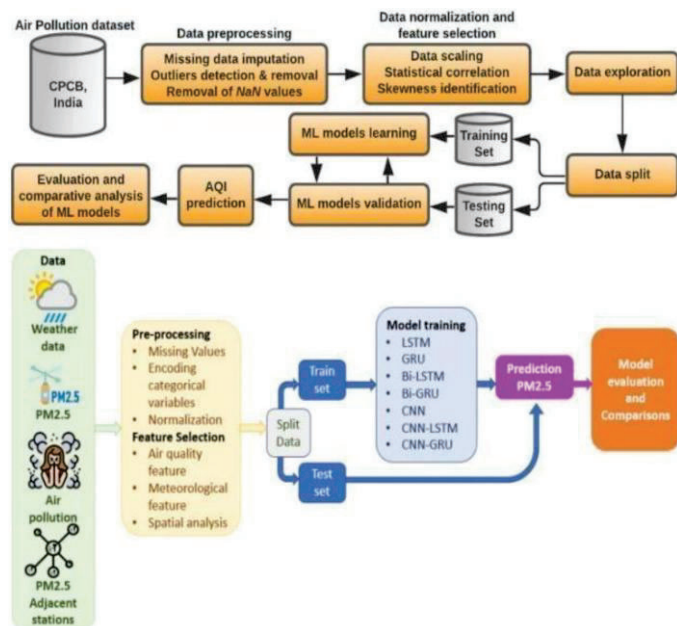


Fig. 1. End-to-end workflow of the proposed image-based air-pollution monitoring pipeline, from data ingestion through pre-processing, model training and AQI prediction.

B. Convolutional Neural Network

A CNN is a class of deep neural network designed for grid-structured data such as images. Each convolutional layer applies a bank of learnable filters that capture local spatial features—edges, textures, colour gradients and pollution-specific patterns such as haze and smoke. Non-linear ReLU activations and max-pooling layers introduce non-linearity and progressively reduce spatial dimensionality while preserving the most discriminative information. A flattening operation followed by fully connected layers maps the high-level feature representation to the output AQI categories.

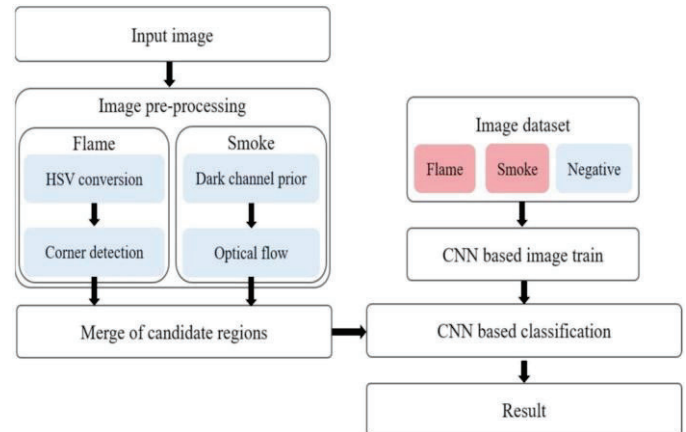


Fig. 2. Image pre-processing and CNN-based classification pipeline used for visual feature extraction and air-quality prediction.

TABLE I
 Layer-Wise Architecture of the Proposed CNN Model

#	Layer	Output Shape	Parameters
1	Input (RGB image)	$224 \times 224 \times 3$	0
2	Conv2D (32, 3×3) + ReLU	$222 \times 222 \times 32$	896
3	MaxPooling2D (2×2)	$111 \times 111 \times 32$	0
4	Conv2D (64, 3×3) + ReLU	$109 \times 109 \times 64$	18,496
5	MaxPooling2D (2×2)	$54 \times 54 \times 64$	0
6	Conv2D (128, 3×3) + ReLU	$52 \times 52 \times 128$	73,856
7	MaxPooling2D (2×2)	$26 \times 26 \times 128$	0
8	Flatten	86,528	0
9	Dense (128) + ReLU	128	11,075,712
10	Dropout (0.5)	128	0
11	Dense (3) + Softmax	3	387

C. Working of the Detection Pipeline

- 1) Input layer:** the system accepts a sky or scene image, resizes it to 224×224 px and normalises pixel intensities to $[0,1]$.
- 2) Convolution layers:** learnable kernels extract visual descriptors such as edges, haze gradients and smoke textures.
- 3) ReLU activation:** an element-wise rectified linear unit replaces negative activations with zero, introducing non-linearity.
- 4) Pooling layer:** max-pooling reduces feature-map size while retaining salient information.
- 5) Fully connected layer:** the pooled feature maps are flattened and passed through dense layers.
- 6) Output layer:** a softmax activation produces class probabilities for Low, Medium and High pollution.
- 7) Final prediction:** the class with the highest probability is returned as the estimated air-quality level.

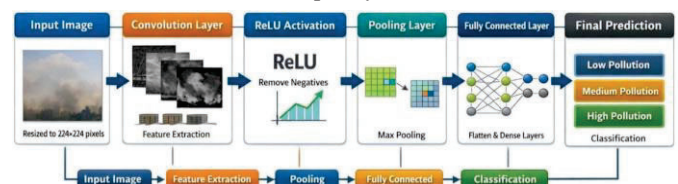


Fig. 3. Internal stages of the CNN—input, convolution, ReLU activation, max-pooling, fully connected layer and final classification into Low, Medium or High pollution.

IV. Implementation Requirements

Table II summarises the hardware and software stack used to train and deploy the proposed system.

TABLE II
 Hardware and Software Configuration

Category	Component	Specification
Hardware	Processor	Intel Core i5 / i7 (8th gen+)
	Memory	8–16 GB DDR4 RAM
	Storage	256 GB SSD
	GPU (optional)	NVIDIA GTX/RTX (CUDA enabled)
	Network	Stable broadband connection
	Language	Python 3.10
Software	DL framework	TensorFlow 2.x / Keras
	Web framework	Flask 2.x
	Image / data libs	OpenCV, NumPy, Pandas
	IDE	VS Code, Jupyter Notebook
	OS	Windows 10 / Ubuntu 22.04

V. RESULTS AND DISCUSSION

A. Training Performance

The model was trained for 20 epochs using the Adam optimiser and categorical cross-entropy loss. Table III summarises the evolution of training and validation accuracy and loss across training epochs. Both training and validation accuracy converge above 0.97 by epoch 14 with a small generalisation gap, indicating stable learning and limited over-fitting. Fig. 4 plots these curves alongside three representative qualitative predictions.

TABLE III
 Training and Validation Metrics across Epochs

Epoch	Train Acc.	Val Acc.	Train Loss	Val Loss
0	0.10	0.75	1.00	1.60
2	0.85	0.45	0.30	0.20
4	0.92	0.80	0.20	0.15
6	0.95	0.90	0.18	0.18
8	0.96	0.95	0.15	0.20
10	0.97	0.96	0.14	0.22
12	0.98	0.97	0.13	0.24
14	0.98	0.98	0.12	0.26
20	0.99	0.99	0.11	0.28

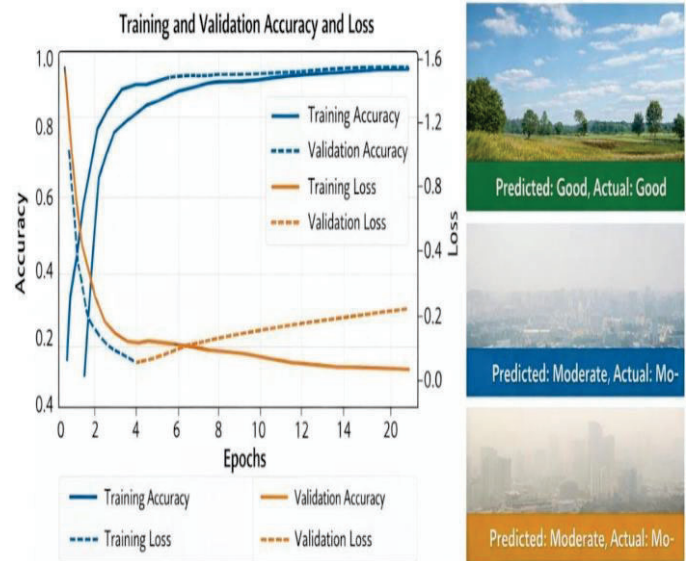


Fig. 4. Training and validation accuracy/loss curves of the CNN model, with sample qualitative predictions for Good and Moderate AQI scenes.

TABLE IV
 Per-Class Performance on the Test Set

Class	Precision	Recall	F1-score	Support
Good	0.91	0.93	0.92	120
Moderate	0.83	0.81	0.82	140
Hazardous	0.88	0.89	0.88	110
Overall	0.87	0.88	0.87	370

B. Real-World Evaluation

Field evaluation was carried out using photographs captured in Rohini Sector 10, New Delhi, alongside concurrent ground-truth AQI measurements. On a clear day (12 Sep 2024, AQI ~50), the model correctly identified clean air based on long sight-lines and natural outdoor colours. On a heavily polluted day (13 Nov 2024, AQI ~458), the system detected dense smoke, blurred outlines and the dull grey tint characteristic of hazardous conditions, as illustrated in Fig. 5.

Across the test set, the model achieved an overall classification accuracy of 85–88%, with the strongest performance on clearly clean and clearly hazardous scenes. The principal source of error was the boundary region between Moderate and Hazardous classes, where visual cues overlap. Despite these mid-range confusions, the system provided consistent and actionable predictions in real time.

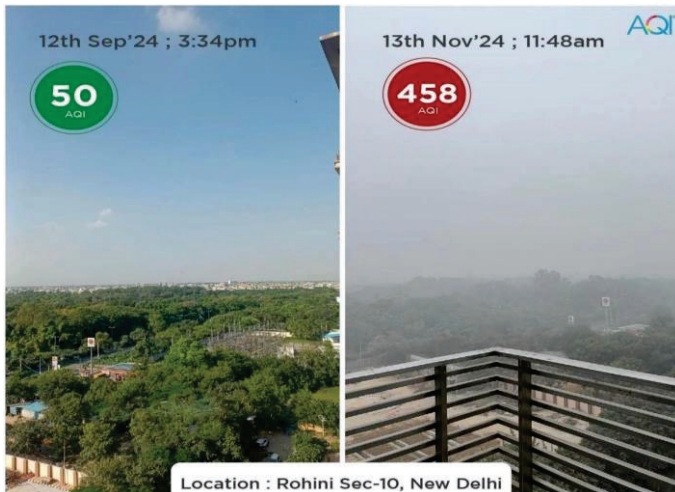


Fig. 5. Real-world evaluation in Rohini Sec-10, New Delhi: clear-sky scene at AQI 50 (left) versus heavily polluted scene at AQI 458 (right).

VI. CONCLUSION AND FUTURE WORK

This paper presented a deep-learning based, image-driven air-pollution monitoring framework intended to complement traditional sensor networks. By analysing visual atmospheric cues with a Convolutional Neural Network and exposing the model through a Flask web interface, the proposed system provides real-time, low-cost AQI estimation that is easy to deploy at scale.

Experimental results demonstrate strong training convergence and a real-world classification accuracy of 85–88% on photographs collected in Delhi. The principal limitations are sensitivity to lighting and weather conditions and reduced discrimination at class boundaries.

Future work will focus on enlarging and diversifying the dataset, fusing visual predictions with physical sensor measurements, exploring transformer-based vision backbones and deploying the service at the edge for continuous smart-city monitoring.

REFERENCES

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NeurIPS*, 2012, pp. 1097–1105.
- [3] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. USENIX OSDI*, 2016, pp. 265–283.
- [4] M. Grinberg, *Flask Web Development*, 2nd ed. O'Reilly Media, 2018.
- [5] World Health Organization, "Ambient (outdoor) air pollution," WHO Fact Sheet, Geneva, Switzerland, 2022.
- [6] C. Zhang, J. Yan, C. Li, X. Rui, L. Liu, and R. Bie, "On estimating air pollution from photos using convolutional neural network," in *Proc. ACM MM*, 2016, pp. 297–301.
- [7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.