



Air quality prediction from images in Indonesia: enhancing model explainability through visual explanation with AQI-net and grad-CAM

Muhammad Labib Alauddin, Novanto Yudistira D and Muhammad Arif Rahman

Departemen Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya, Malang, Indonesia Corresponding author: Novanto Yudistira; Email: yudistira@ub.ac.id

Received: 22 June 2025; Revised: 22 June 2025; Accepted: 23 June 2025

Keywords: Air Quality Index; deep learning; digital image; image processing; visual explanation

Abstract

Good air quality is a critical determinant of public health, influencing life expectancy, respiratory health, work productivity, and the prevention of chronic diseases. This study presents a novel approach to classifying the Air Quality Index (AQI) using deep learning techniques, specifically convolutional neural networks (CNNs). We collected and curated a dataset comprising 11,000 digital images from three distinct regions in Indonesia—Jakarta, Malang, and Semarang—ensuring uniformity through standardized acquisition settings. The images were categorized into four air quality classes: good, moderate, unhealthy for sensitive groups, and unhealthy. We designed and implemented a CNN architecture optimized for AQI classification. The model achieved an impressive accuracy of 99.81% using K-fold cross-validation. In addition, the model's interpretative capabilities were examined using techniques such as Grad-CAM, providing valuable insights into how the CNN identifies and classifies air quality conditions based on image features. These findings underscore the effectiveness of CNNs for AQI classification and highlight the potential for future work to incorporate a more diverse set of digital images captured from various perspectives to enhance dataset complexity and model robustness. The dataset is publicly accessible at https://doi.org/ 10.5281/zenodo.15727522.

Impact Statement

This article addresses a pressing global issue: the assessment of air quality, a vital determinant of public health and environmental management. By introducing AQI-Net, a specialized Convolutional Neural Network model, this study pioneers the use of deep learning to classify the Air Quality Index (AQI) from digital images. This collaborative effort between experts in computer vision, environmental data analysis, and software development guarantees a multidisciplinary perspective. The study not only provides a comprehensive and publicly accessible dataset but also enhances model explainability through Grad-CAM, offering valuable insights into the decision-making process of Artificial Intelligence models for broader scientific and public health applications.



👔 😝 This application paper was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

1. Introduction

Air is a mixture of gases, primarily nitrogen, oxygen, and carbon dioxide, that are essential for the survival of living organisms. Oxygen, in particular, is vital for respiration, a process that sustains life. Consequently, the quality of air is directly linked to public health. Good air quality supports longer life expectancy, healthier respiratory systems, improved work productivity, and a reduced incidence of chronic diseases. However, the natural composition of air can be disrupted by the introduction of harmful substances, leading to air pollution. This disruption often results from various human activities, such as industrial emissions, vehicle exhaust, cigarette smoke, deforestation, and large-scale agricultural practices like the burning of crop residues (Maharani and Aryanta, 2023). As air pollution intensifies, it poses significant risks to public health, making the assessment and monitoring of air quality increasingly important.

The assessment of air quality typically involves measuring a set of parameters, including the levels of ozone (O₃), carbon monoxide (CO), sulfur dioxide (SO₂), nitrogen dioxide (NO₂), particulate matter (PM), and other pollutants (Huboyo et al., 2020). These measurements are used to determine the Air Quality Index (AQI), a standardized index that categorizes air quality into different classes, each reflecting the associated health risks of various pollution levels. Traditionally, the AQI is determined using sophisticated sensors that detect pollutant concentrations in the air (Yu et al., 2018). While accurate, these sensors are expensive and often limited to major urban areas, restricting their accessibility. To overcome these limitations, alternative approaches to estimating the AQI have been explored, including the use of digital images. By capturing the visual appearance of the sky or surroundings, these images can be analyzed using deep learning techniques, specifically convolutional neural networks (CNNs). This method offers a cost-effective and accessible solution for monitoring air quality, particularly in regions where sensor deployment is not feasible.

Deep learning, a branch of machine learning, is particularly effective for tasks that involve large datasets and complex patterns. Unlike traditional machine learning models, which often require manual feature extraction, deep learning architectures, such as CNNs, are designed to automatically discover intricate patterns through hierarchical feature extraction. This makes CNNs highly suitable for analyzing visual data, including digital images used for air quality assessment.

A CNN is composed of several types of layers, each serving a distinct function. The three primary types of layers are convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply kernels (or filters) to the input image, performing operations that detect various features, such as edges, textures, and shapes. These features are then transformed into feature maps that capture the essential characteristics of the image. The pooling layers downsample these feature maps, reducing their spatial dimensions while retaining the most important information, making the model more computationally efficient. Finally, the fully connected layers process the extracted features to produce the final output, such as a classification label or probability distribution (Popescu et al., 2009).

One practical application of CNNs is in determining the AQI based on digital images. The AQI is a globally recognized index used to communicate the current or forecasted level of air pollution. It is typically divided into six categories: good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy, and hazardous (Agista et al., 2020). Each category has specific characteristics and health implications. For example, the "good" category (green) with a PM2.5 value of 0–12 indicates ideal conditions for outdoor activities and suggests that people can enjoy fresh air by opening windows. In contrast, the "hazardous" category (maroon) reflects severe air pollution levels that can lead to serious health consequences, including an increased risk of cardiovascular disease (Zhao et al., 2020).

However, existing image-based AQI models primarily focus on daytime imagery under well-lit conditions, limiting their applicability to 24-h monitoring.

By leveraging CNNs to classify digital images according to the AQI, we can develop tools that provide a cost-effective means of public health monitoring and environmental management, particularly in areas lacking extensive sensor networks.

CNNs have already demonstrated success in various applications beyond image classification, including voice emotion recognition, food composition analysis, and sentiment analysis (Yu et al., 2022). This makes them a promising approach for addressing the challenges of air quality assessment.

2. Dataset

Choosing suitable locations for image capture in Indonesia requires careful consideration due to the country's vast size and geographical diversity. Indonesia spans an area of ~1.905 million km², making it challenging to collect data that adequately represents the entire nation. In addition, validating the AQI labels used in classifying digital images necessitates reliance on pollutant detection sensors, which are not uniformly distributed across the country. As a result, data collection was concentrated in regions with reliable AQI observations and sufficient sensor coverage.

Given Indonesia's size and diversity, we selected regions that are representative of the country's various environmental conditions. These regions were chosen based on factors such as population density, industrial activity, geographical features, and the availability of pollutant detection sensors. For instance, Sumatra Island is prone to forest fires during the dry season (Yusuf et al., 2019), while Kalimantan Island is experiencing rapid industrial growth, largely due to the expansion of oil palm plantations (Huda et al., 2021). Java Island, the most populous and industrialized island in Indonesia, was chosen as the primary focus of this study due to its extensive sensor network and diverse environmental conditions (Mardiansjah and Rahayu, 2019).

To ensure comprehensive coverage, Java Island was divided into three regions: western Java, represented by Jakarta; central Java, represented by Semarang; and eastern Java, represented by Malang. Jakarta, the capital city, was selected not only for its high population density but also for its status as an economic and political hub, which contributes to its varied air quality challenges.

The datasets used in this study were collected from Central Jakarta, Semarang, and Malang. Data collection was conducted using the POCO X3 Pro device, with images captured throughout March and April. The air quality at the time of image capture was verified using the IQAir website at https://www.iqair.com/id/, which provides real-time AQI data. These images were collected specifically for this study (not sourced from any existing online database), ensuring that each image's label corresponds to the actual measured AQI at the time of capture. This dataset serves as a valuable resource for training CNN models to classify air quality based on visual data, offering a cost-effective alternative to traditional sensor-based methods. By capturing a wide range of environmental conditions across multiple locations and times, this dataset provides a robust foundation for developing models that can generalize well to new, unseen data.

Based on Table 1, we can see that the AQI can be categorized based on a predefined range of values. Therefore, we labeled our dataset according to this reference table, ensuring that each image's AQI value falls within the correct category (Attaallah and Khan, 2022). The dataset taken from the city on Java Island was only able to capture four classes, which were labeled as "good," "moderate," "unhealthy for some people," and "unhealthy."

Figure 1 displays the sample images representing each of the four AQI classes. Each image taken at the same place has three different angles to add variety and complexity to the data collected. Another reason for requiring multiple angles is to maximize the objects contained in the image so that the model does not misrecognize the pattern.

| AQI color | Range value | Description |
|-----------|----------------|--------------------------------|
| Green | 0–50 | Good |
| Yellow | 51–100 | Moderate |
| Orange | 101–150 | Unhealthy for sensitive groups |
| Red | 151–200 | Unhealthy |
| Purple | 201–300 | Very unhealthy |
| Maroon | 301 and higher | Hazardous |

Table 1. Air Quality Index class (Li et al., 2017)



Figure 1. Examples of datasets collected clockwise, ranging from good, moderate, unhealthy for some people, and unhealthy air quality.

To ensure that we have enough variety of the data gathered, image collection was conducted at three distinct locations across three different cities. This variation in data is necessary to represent all relevant aspects of the observed scene (Barbedo, 2018).

Figure 2 illustrates the locations in Jakarta, Semarang, and Malang where images were captured (red dots mark the AQI sensor points). These locations were identified using the AQI dashboard feature provided by IQAir (IQAir). The data collection strategy was to take photographs in the vicinity of each AQI sensor so that the measured AQI corresponds directly to the scene captured in the image.

3. Data acquisition

Images were taken by dividing the shooting times into three sessions: morning, afternoon, and evening. The morning shooting session was conducted from 8 to 11 A.M., the afternoon session from 12 to 2 P.M., and the evening session from 3 to 5 P.M. Thus, no images were captured after 5 P.M. (i.e., under nighttime or low-light conditions). After that, there is the image capture process, which is used to capture images and label them based on the AQI detection results around the image capture location. Finally, there is the image-cleaning phase. This phase is used to clean the image from unnecessary noise, such as foreign objects appearing during image capture, poor capture results, or blur (Pal and Sudeep, 2016).

4. AQI-Net

After completing the data collection, the digital image data are processed using a deep learning architecture known as CNNs. In this article, the modified CNN is referred to as AQI-Net. The AQI-Net architecture comprises three convolutional blocks. The first and second blocks each consist of one convolutional layer, one activation function, and one max pooling layer. The final block includes a linear



Figure 2. Shooting locations in Jakarta, Semarang, and Malang, with red dots on the picture indicating the exact shooting points and numbers representing the AQI sensors.

layer that serves as the classification layer. Designing the architecture involves careful consideration of the design complexity, training time, and accuracy achieved during testing. After arranging and researching the necessary layers, the following architecture was developed.

As shown in Table 2, the complete architecture of the modified CNN, named AQI-Net, used for AQI classification is depicted. This architecture will be trained and tested to assess its performance in classifying data into four different AQI categories. The model employs an input shape of 224×224 with three channels, which is processed through the first convolutional block for spatial reduction. This block summarizes the information in the digital image, condensing multiple pieces of information into a single representation. In the first convolutional layer, a 5×5 kernel with a stride of 1 is used. Spatial reduction also helps accelerate training time. The max-pooling operation in the convolutional Block 1 uses a 2×2 kernel with a stride of 2.

Subsequently, the data progress to the second convolutional block for additional spatial reduction. This block mirrors the structure of the first, consisting of a convolutional layer with a 5×5 kernel and a stride of 1, and a max-pooling layer with a 2×2 kernel and a stride of 2. Finally, the third convolutional block includes a linear (or flattening) layer that converts the data into a one-dimensional vector. At this stage, the output from the second convolutional block is a feature map of size $53 \times 53 \times N$ (where N is the number of feature maps in Block 2). The flattening layer converts this into a one-dimensional vector of $53 \times 53 \times N = 140,145$ features. These features are then reduced to 300 via a fully connected layer with a ReLU (Rectified Linear Unit) activation function. In the final layer, these 300 features are processed through another fully connected layer, which classifies the data into predefined categories, completing the supervised learning process. Having established the AQI-Net architecture and prepared the dataset, we next trained the model and evaluated its performance, as presented in the following section.

| Layer (type) | Output shape | Kernel size/ stride | Details |
|------------------------|-------------------------|------------------------|---|
| Input layer | (224, 224, 3) | _ | Input shape of $224 \times 224 \times 3$ |
| Conv2D (Block 1) | (220, 220, numChannels) | 5 × 5/1 | Spatial reduction, 5 × 5 kernel, stride 1 |
| MaxPooling2D (Block 1) | (110, 110, numChannels) | 2 × 2/2 | MaxPooling with 2 × 2 kernel, stride 2 |
| Conv2D (Block 2) | (106, 106, numChannels) | 5 × 5/1 | Spatial reduction, 5 × 5 kernel, stride 1 |
| MaxPooling2D (Block 2) | (53, 53, numChannels) | 2 × 2/2 | MaxPooling with 2 × 2 kernel, stride 2 |
| Flatten (Block 3) | (140, 145) | _ | Flattening layer |
| Dense (Block 3) | (300) | _ | Fully connected, ReLU activation |
| Output layer (Dense) | (4) | - | Fully connected, four output classes |

Table 2. Architecture of the proposed AQI-Net

5. Results

We evaluated the model using a fivefold cross-validation approach (with K=5) on the combined image dataset from Jakarta, Semarang, and Malang. The images were randomly divided into five equal folds with approximately uniform class distributions. In each iteration of cross-validation, four folds (80% of the data) were used for training, and the remaining one fold (20%) was used for validation. This process was repeated five times so that each fold was used exactly once as the validation set. Using this strategy, the proposed AQI-Net achieved an average validation accuracy of 99.81% (with the highest single-fold accuracy reaching 99.97%) across the five folds, demonstrating the model's strong generalization performance. All accuracy values reported in this section correspond to validation results from the cross-validation.

A comparative analysis of various architectures offers valuable insights into the suitability of the collected dataset for classification tasks. The architectures evaluated for performance comparison include ResNet50 (He et al., 2015), VGG16 (Simonyan and Zisserman, 2015), ColorNet (Zhang et al., 2017), and the proposed AQI-Net. This comparative study aims to assess whether AQI-Net achieves comparable or superior performance relative to the benchmark architectures. ResNet50 and VGG16 were selected as representative deep CNN models due to their proven performance in image classification, providing strong baselines for comparison. We also included a colorization-based network (ColorNet) to examine whether modeling color distributions in images can aid air quality classification, since atmospheric color (e.g., haziness or sky tint) can be an indicator of pollution levels. The performance metrics for these models are summarized in Table 3.

Architecture Pretrained? Accuracy Training time (s) Total parameters ResNet50 100% Yes (ImageNet) 2486.3 23,516,228 VGG16 100% Yes (ImageNet) 10263.7 134,276,932 ColorNet 99.88% Yes (ImageNet) 5260.1 423,085,767 AQI-Net 99.97% No (Scratch) 2068.6 42,163,074

Table 3. Performance comparison of models

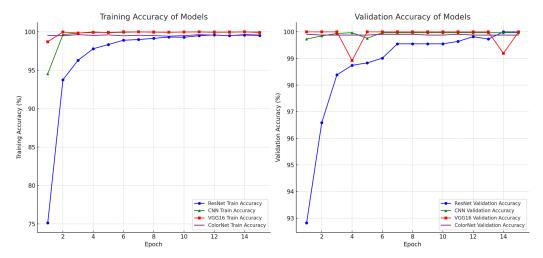


Figure 3. Comparison of several architectures on the Indonesian dataset.

As shown in Table 3, all models demonstrate exceptional performance on the validation data, with ResNet50, VGG16, and AQI-Net each achieving near-100% validation accuracy. ColorNet's accuracy is slightly lower but still excellent. AQI-Net is particularly noteworthy for its efficiency, achieving high accuracy with the shortest training time. In contrast, while VGG16 is highly accurate, it requires significantly more training time compared to the other models, which may be a consideration when computational resources are limited. The table also lists each model's total number of parameters, highlighting differences in model complexity. We observe that VGG16 and ColorNet have substantially more parameters (~134 million and 423 million, respectively) compared to AQI-Net (42 million) and ResNet50 (23 million). AQI-Net's parameter count, while much lower than those of VGG16 and ColorNet, is still relatively high—this is primarily due to its large fully connected layer (flattening roughly 140k features into 300 nodes), which contributes the majority of its 42 million parameters. We have double-checked these values for accuracy.

Based on Figure 3, we can evaluate the performance of the architectures trained using the Indonesian dataset. Initially, ColorNet exhibits a significant gap between training and validation accuracies, indicating overfitting, where training accuracy surpasses validation accuracy. However, the model stabilizes in subsequent epochs and eventually converges. In contrast, AQI-Net demonstrates a minimal difference between validation and training accuracy and loss, reflecting stable performance and effective recognition of the dataset. VGG16 and ResNet50 also show stable performance, although ResNet50 does not initially achieve the best results compared to AQI-Net and VGG16.

The graph displays the progression of training accuracy for each model over 15 epochs. All models show a steady increase in accuracy, with some models, such as AQI-Net and ColorNet, reaching near-perfect accuracy by the end of the training period. Validation accuracy is also plotted over the same epochs, with all models maintaining high validation accuracy. Notably, ResNet50 and VGG16 achieve 100% validation accuracy, indicating excellent generalization on the validation dataset (Novak et al., 2018).

5.1. AQI-Net model explanation via Grad-CAM

Grad-CAM visualizes which parts of the image the model focuses on to make its classification decision, providing insight into the model's reasoning (Selvaraju et al., 2017). In the test results using Grad-CAM for AQI-Net, we can observe how the model determines the class of a digital image. For instance, when the model classifies an image as belonging to the "good" class, Grad-CAM highlights the regions of the image associated with features relevant to the "good" label. This visualization demonstrates that the model's

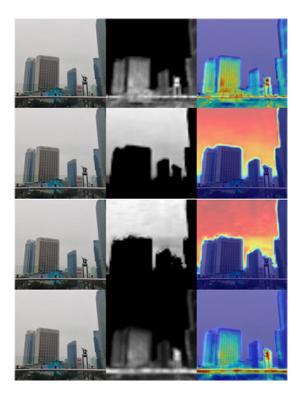


Figure 4. Testing the AQI-Net model with Grad-CAM on an image labeled "Unhealthy for Some": Each row in the figure corresponds to a different target class from the dataset, starting from the top: "good," "moderate," "unhealthy for some," and "unhealthy."

classification decision is based on these significant structures or features. When we track Grad-CAM visualizations using the true class label as the target, the highlighted regions tend to align intuitively with features a human might also consider relevant, such as the sky in the context of air quality. However, when Grad-CAM is computed using an incorrect or nontrue class label, the resulting heatmaps often focus on less meaningful or even unrelated parts of the image, making them less sensible from a human interpretability standpoint. This contrast can serve as a qualitative sanity check on the model's internal reasoning.

Figure 4 presents Grad-CAM visualizations that elucidate the AQI-Net model's interpretative focus across different air quality labels: "good," "moderate," "unhealthy for some," and "unhealthy." For the image labeled "unhealthy for some," the model predominantly highlights the sky region, which corresponds to human perceptual tendencies, where the sky is often indicative of air quality. In contrast, the heatmap for the "good" label reveals a focus on structural elements, such as buildings, which is less intuitive, since we generally associate good air quality with clear skies rather than man-made structures. This mismatch further illustrates how Grad-CAM responses for nontrue classes may not always make sense from a human interpretability standpoint.

6. Conclusion

This research demonstrates the effectiveness of CNNs for classifying air quality into four categories: good, moderate, unhealthy for sensitive groups, and unhealthy. Using real-time AQI data from Jakarta, Malang, and Semarang, the proposed AQI-Net model achieved near-perfect accuracy in Jakarta and Semarang, with slightly lower performance in Malang due to data variability. Compared to ResNet50,

VGG16, and ColorNet, AQI-Net stands out for its efficiency, requiring significantly less training time while maintaining high accuracy.

Grad-CAM analysis revealed that AQI-Net focuses on structural elements like buildings and skies for classification, although its reliance on less intuitive features (e.g., buildings) suggests room for improvement. Overall, these visualizations provide useful explanations of the model's focus (e.g., highlighting the sky region for poorer air quality). However, they largely confirm the expected cues rather than uncovering fundamentally new insights into the model's decisions. Despite this, AQI-Net's stable performance and fast convergence make it a robust and efficient solution for air quality classification.

AQI-Net offers a balance of accuracy and efficiency, making it a valuable tool for environmental monitoring. However, the current model is limited to daytime scenarios because no nighttime images were included in the training. Future work should address this by incorporating low-light (evening/night) images or applying image enhancement for low-light conditions, thereby extending the approach to 24-h monitoring. In addition, future studies should focus on expanding the dataset diversity (for instance, by including nighttime imagery) and further improving model interpretability to better align the system with the human perception of air quality.

Author contribution. Data curation: M.L.A.; Project administration: M.A.R.; Conceptualization: N.Y.

Competing interests. The authors declare none.

Data availability statement. The code and data used in this study have been archived on Zenodo and are publicly available at https://doi.org/10.5281/zenodo.15727522 (Alauddin and Yudistira, 2025).

Ethics statement. The authors confirm that all data were collected in accordance with the applicable laws and regulations of Indonesia. No human or animal subjects were involved in this study.

Funding statement. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Provenance statement. This article was accepted into the Climate Informatics 2025 (CI2025) Conference. It has been published in Environmental Data Science on the strength of the CI2025 review process.

References

Agista P, Gusdini N and Maharani M (2020) Analisis kualitas udara dengan indeks standar pencemar udara (ispu) dan sebaran kadar polutannya di provinsi dki Jakarta. Sustainable Environmental and Optimizing Industry Journal 2, 39–57.

Alauddin ML and Yudistira N (2025). Air quality index dataset in Indonesia for classification. Zenodo. https://doi.org/10.5281/zenodo.15727522

Attaallah A and Khan RA (2022) Smotednn: A novel model for air pollution forecasting and aqi classification. Computers, Materials & Continua 71, 1403–1425.

Barbedo JGA (2018) Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture 153*, 46–53.

He K, Zhang X, Ren S and Sun J (2015) Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770–778.

Huboyo HS, **Hadiwidodo M and Nurihsan M** (2020) Konsentrasi anion di udara ambien dan analisis lintasan balik sumber polutan di Kota Semarang. *Jurnal Serambi Engineering* 5, 10.

Huda IU, Karsudjono AJ and Darmawan R (2021) Analisis bonus demografi terhadap pertumbuhan ekonomi di provinsi Kalimantan Selatan. *Al-KALAM: JURNAL KOMUNIKASI, BISNIS DAN MANAJEMEN 8*(2), 1–21.

IQAir Yang Pertama dalam Kualitas Udara. iqair.com. Available at https://www.iqair.com/id/ (accessed 06 August 2024).

Li Y, Tang Y, Fan Z, Zhou H and Yang Z (2017) Assessment and comparison of three different air quality indices in China. Environmental Engineering Research 23, 21–27.

Maharani S and Aryanta WR (2023) Dampak buruk polusi udara bagi kesehatan dan cara meminimalkan risikonya. *Jurnal Ecocentrism* 3, 47–58.

Mardiansjah FH and Rahayu P (2019) Urbanisasi dan pertumbuhan Kota-Kota di Indonesia: Suatu perbandingan antar-wilayah makro Indonesia. *Jurnal Pengembangan Kota* 7(1), 91–110.

Novak R, Bahri Y, Abolafia DA, Pennington J and Sohl-Dickstein J (2018) Sensitivity and generalization in neural networks: An empirical study. arXiv preprint arXiv:1802.08760.

Pal KK and Sudeep KS (2016) Preprocessing for image classification by convolutional neural networks. In 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 1778–1781.

- Popescu M-C, Balas VE, Perescu-Popescu L and Mastorakis N (2009) Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems 8(7), 579–588.
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D and Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, pp. 618–626.
- Simonyan K and Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556.
- Yu H, Cheng X, Chen C, Heidari AA, Liu J, Cai Z and Chen H (2022) Apple leaf disease recognition method with improved residual network. Multimedia Tools and Applications 81, 7759–7782.
- Yu T, Wang W, Ciren P and Sun R (2018) An assessment of air-quality monitoring station locations based on satellite observations. *International Journal of Remote Sensing* 39(20), 6463–6478.
- Yusuf A, Hapsoh H, Siregar SH and Nurrochmat DR (2019) Analisis kebakaran hutan dan Lahan di provinsi Riau. *Dinamika Lingkungan Indonesia* 6(2), 67–84.
- Zhang R, Zhu J-Y, Isola P, Geng X, Lin AS, Yu T and Efros AA (2017) Real-time user-guided image colorization with learned deep priors. ACM Transactions on Graphics (TOG) 9(4), 1–11.
- Zhao B, Johnston FH, Salimi F, Kurabayashi M and Negishi K (2020) Short-term exposure to ambient fine particulate matter and out-of-hospital cardiac arrest: A nationwide case-crossover study in Japan. *The Lancet Planetary Health* 4(1), e15–e23.

Cite this article: Alauddin ML, Yudistira N and Rahman MA (2025). Air quality prediction from images in Indonesia: enhancing model explainability through visual explanation with AQI-net and grad-CAM. *Environmental Data Science*, 4: e42. doi:10.1017/eds.2025.10011