DECODE: JURNAL PENDIDIKAN TEKNOLOGI INFORMASI
P-ISSN: 2775-2984 | E-ISSN: 2775-1813

Vol. 5, No 3, November 2025, Hal: 1293-1311
Doi: http://dx.doi.org/10.51454/decode.v5i3.1412
https://journal.umkendari.ac.id/index.php/decode

Optimization of Sky Image Enhancement to the Performance of Air Quality Classification Model Using Mobilenet V3

Wempy Aditya Wiryawan[1*], Christian Sri Kusuma Aditya[1]

[1]Informatics Study Program, Universitas Muhammadiyah Malang, Indonesia.

Artikel Info

Keywords:

Air Quality;
Deep Learning;
Image Preprocessing;
Mo-bileNetV3;
Sky Image.

Article History:

Submitted: September 19, 2025
Accepted: November 30, 2025
Published: November 30, 2025

Abstract: Air quality monitoring based on sky imagery is an alternative solution to the limitations of conventional monitoring tools. However, raw sky imagery often has low contrast and visual noise that can hinder the performance of deep learning models in recognizing atmosphere patterns. Therefore, this study aims to develop an efficient and accurate Air Quality Index (AQI) estimation model using MobileNetV3. The novelty of this research lies in the systematic investigation of image enhancement techniques using the public Air Pollution Image Dataset from India and Nepal, an aspect that has not been deeply explored in previous studies. To achieve this goal, four pre-processing techniques—Histogram Equalization (HE), Contrast Limited Adaptive Histogram Equalization (CLAHE), Gamma Correction, and Contrast Stretching—were applied and compared. The results consistently show that Histogram Equalization (HE) is the most superior technique for regression tasks with an R^2 value of 0.913, as well as for classification tasks with the highest category accuracy of 0.690. These findings significantly outperform models without pre-processing, confirming that HE is the most recommended technique for maximizing accuracy on resource-constrained devices.

Corresponding Author:
Wempy Aditya Wiryawan
Email: adityawempy@gmail.com

INTRODUCTION

Air quality has an important role in public health and environmental sustainability. Fine dust particles such as PM10 can enter the human respiratory system and cause a variety of serious health problems, including lung and cardiovascular diseases. Data from Global Burden Diseases 2019 shows that in Indonesia, of the 10 diseases with the most cases per 100,000 population, 4 of them are respiratory diseases, which contribute to the high mortality rate (Rokom, 2023) Air pollution also increases the risk of cardiovascular diseases, including heart disease and stroke. Air quality monitoring is important, especially in urban areas with high emission levels. Although the Ministry of Environment and Forestry (MoEF) has developed an *Air Quality Monitoring System* (AQMS) for *real-time* monitoring, constraints such as installation costs and device limitations in remote areas are still major challenges. Alternatively, digital image processing offers a more affordable and accessible solution in air quality monitoring. Sky image analysis using a simple device such as a mobile camera can be a practical approach to monitor air quality in real-time. With the advancement of (Amahoru et al., 2024). *Machine Learning* (ML) and *Deep Learning* (DL) technologies, particularly *Convolutional Neural Networks* (CNNs), the analysis and classification of air quality can be carried out automatically based on the visual characteristics of the sky imagery.(Azlansyah et al., 2019)

Several previous studies have explored the use of image processing and deep learning in air quality monitoring. Samsami in his research used the Local Binary Patterns (LBP) method and Gabor filters for feature extraction, with the k-Nearest Neighbor (k-NN) classification algorithm, and managed to achieve an accuracy of 82.02% (Samsami et al., 2019). Vahdatpour compares shallow classifiers to CNNs, where shallow classifiers perform better on small datasets, whereas CNNs have the potential to be deployed on a wider scale with optimized architectures. In the study conducted by Enggari using the color histogram method, which managed to achieve 90% accuracy in classifying air quality into three categories: good, moderate, and unhealthy (Mukundan et al., 2022; Enggari et al., 2024). However, most of these studies focused predominantly on model architectures or feature extraction methods, often utilizing raw images without adequate quality improvement stages. This presents a significant research gap, as the issues of low contrast and visual noise in sky imagery remain largely unaddressed, thereby hindering the model's ability to optimally recognize atmospheric patterns.

Previous studies have shown promising results in the use of sky imagery for air quality monitoring, but there are still limitations that need to be overcome, one of which is the lack of optimization of image preprocessing techniques. Some studies tend to use *raw images* without improving image quality first. This leads to *visual noise*, low contrast, and uneven lighting, which makes it difficult for the model to recognize atmospheric patterns optimally. For example, in some datasets, misclassification still occurs due to cloud texture or blurred sky color gradation due to (Bismi & Qomaruddin, 2023)*underexposed* lighting conditions or thin fog. (Kamil, 2022)

To overcome these limitations, this study applies and compares four image quality improvement techniques, namely Histogram Equalization (HE), Contrast Limited Adaptive Histogram Equalization (CLAHE), Gamma Correction, and Contrast Stretching. These four techniques were chosen because they represent different fundamental approaches to image enhancement: HE and Contrast Stretching as a global method, CLAHE as a local adaptive method, and Gamma Correction which focuses on luminance adjustment. This selection allows for a comprehensive comparative analysis to determine which approach is most effective in extracting atmospheric features for deep learning models, aiming to improve visual quality, clarify atmospheric patterns, and support a more optimal feature extraction process.

Various CNN architectures have been explored for image-based air quality estimation tasks. For example, research by Utomo, 2023. compare several pre-trained models including VGG16, ResNet50, and MobileNetV3 for similar tasks. In their experiments, the VGG16 showed the highest accuracy and was eventually chosen as the base model. Nonetheless, the study specifically chose to use MobileNetV3 as the main model. The main reason for this choice is that MobileNetV3 is designed to have high computing efficiency, making it more suitable for implementation on devices with limited resources(Jochsen & Handhayani, 2024). The dataset used was obtained from secondary sources,

namely a collection of sky images with air quality index (AQI) information that had been validated and used in previous research [13]. Model evaluation was carried out using classification metrics such as *Confusion Matrix*, *Precision*, *Recall*, *F1-Score*, as well as regression metrics such as *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), and *R-squared* (R^2) to measure the accuracy of the prediction of the AQI value. In addition, an evaluation of the accuracy of category predictions based on preprocessing results was also carried out to compare the effectiveness of each technique in supporting classification performance.

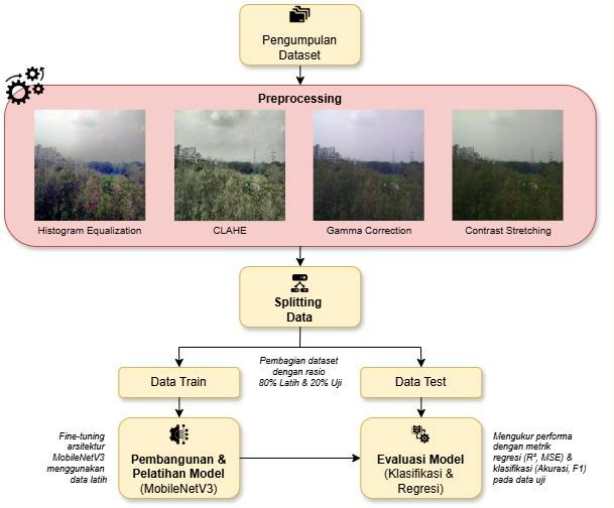
The urgency of the development of the model proposed in this study is reinforced by the extreme air quality conditions reported in various countries, including areas with high pollution levels and rapid urbanization. Accordingly, the main contribution of this study lies in a systematic comparative analysis of the influence of four image *preprocessing techniques* on the performance of the MobileNetV3 model. The study specifically identifies which techniques are most effective in improving the accuracy of image-based air quality estimation, so that it can provide practical guidance for the development of similar systems in the future.

Various environmental studies and reports position India and Nepal as representations of air quality conditions that reflect global challenges, especially in regions with rapid urbanization and high density. Cities such as Delhi and Kathmandu consistently record high levels of air pollution, with the air quality index (AQI) category often being in unhealthy to hazardous status. For example, in November 2024, Delhi recorded an AQI of 491, which falls into the (Dede et al., 2020; Septiyana et al., 2023)*severe plus* category, indicating extremely hazardous air conditions for health. Cases like these show the urgency of developing air quality monitoring systems that are more adaptive, efficient, and widely applicable in different regions with similar atmospheric characteristics.

METHOD

Research Methods

This study applied the Convolutional Neural Network (CNN) architecture of MobileNetV3 to estimate and classify air quality based on sky imagery. The research flow began with the collection of datasets from the Kaggle platform, followed by Exploratory Data Analysis (EDA). Four image quality improvement techniques, namely Histogram Equalization (HE), and CLAHE, Gamma Correction, Contrast Stretching were applied separately. The processed images are then processed through Bilateral Filtering for noise reduction and resized according to MobileNetV3 input. The pre-processed dataset is divided into training and validation data, and then trained using MobileNetV3 with fine-tuning at the initial ImageNet weight. Evaluation is conducted with classification and regression metrics to determine the best pre-processing technique. The research flow chart can be seen in Figure 1.



Picture 1. Research Flow Diagram

Experimental Setup

Experimental Setup All computational experiments and model training in this study were executed using the Kaggle cloud computing platform (Kaggle Kernels). To accelerate the deep learning training process, the environment was configured with hardware specifications consisting of dual NVIDIA Tesla T4 GPUs, a 4-core CPU, and 29 GB of RAM. On the software side, the implementation was built using the Python programming language, utilizing the TensorFlow/Keras framework for the MobileNetV3 architecture and the OpenCV library for executing image pre-processing algorithms.

Dataset Collection

In this study, the dataset used was obtained from a secondary source, namely the Air Pollution Image Dataset from India and Nepal, available on Kaggle (Rouniyar, 2024). This dataset was selected based on its validity established in previous research, such as the Eff-AQI study, which demonstrated its reliability for air pollution estimation tasks (Utomo et al., 2023). The dataset consists of a total of 12,240 images resized to a uniform resolution of 224×224 pixels. Each image is categorized based on the Air Quality Index (AQI) into six classes with each range of AQI values shown in Table 1.

Table 1. Distribution of AQI Value Ranges for Each Air Quality Level

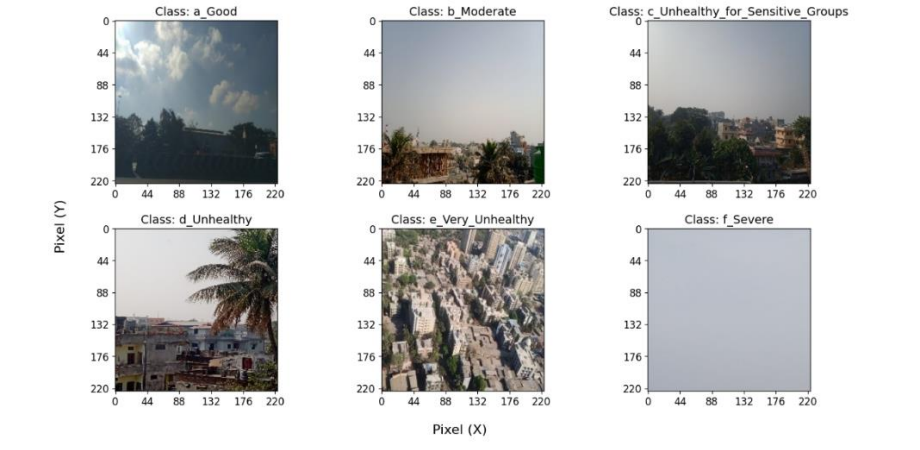
Level	AQI value
Good	0-50
Moderate	51-100
Unhealthy for Sensitive Groups	101-150
Unhealthy	151-200
Very Unhealthy	201-300
Severe	301-500

The dataset structure is divided into two main directories, namely Combined_Dataset and Country_wise_Dataset. The Combined_Dataset directory contains two subfolders: All_img, which includes the entire image, and IND_and_NEP, which contains images that have been grouped by AQI class. There is also a CSV file containing metadata, such as location, file name, date and time of shooting, AQI value, and concentration of key pollutants (PM2.5, PM10, O_3, CO, SO_2, NO_2). Meanwhile, the Country_wise_Dataset directory groups images by country (India and Nepal), with a city subfolder as well as a metadata CSV file, respectively. The initial stage in data exploration is to understand the characteristics of the dataset used. This dataset consists of a total of 12,240 images categorized by AQI. The distribution of the number of images for each AQI category is shown in Table 2.

Table 2. Distribution of Number of Images for Each AQI Class

Level	Number of Images
Good	1.541
Moderate	1.573
Unhealthy for Sensitive Groups	2.863
Unhealthy	2.622
Very Unhealthy	2.194
Severe	1.447

In addition, this dataset includes information from various locations, with varying amounts of images, namely 10,069 images from India and 2,171 images from Nepal. Sample imagery datasets for each level/class of AQI can be seen in the visualization of Figure 2.



Picture 2. Image visualization of each AQI class

Statistical analysis was also carried out on the parameters of air pollutants in the dataset, namely the AQI value. The average values and distribution of each pollution parameter are shown in Table 3.

Table 3. Descriptive Statistics for Air Pollutant Values

	AQI
count	12.240.000.000
mean	167.626.797
Std	102.818.213
min	15.000.000
25%	97.000.000
50%	152.000.000
75%	230.000.000
Max	450.000.000

To ensure objective and reproducible model evaluation, the data splitting scheme was performed using a random shuffle with a ratio of 80% for the training set, 10% for the validation set, and 10% for the testing set. This partition was executed after the application of each pre-processing technique to ensure consistent data distribution across experiments.

Regarding data augmentation and processing, this study focuses on the comparative analysis of image enhancement techniques. Consequently, four distinct dataset variations were constructed by separately applying Histogram Equalization (HE), CLAHE, Gamma Correction, and Contrast Stretching. All images were subsequently normalized to a value range of [-1, 1] to meet the input requirements of the MobileNetV3 architecture. No additional geometric augmentations (such as rotation or flipping) were applied during this comparative phase to isolate the pure effect of the image enhancement techniques on model performance.

Image Preprocessing

The preprocessing stage is carried out to improve the quality of the image before it is used in *deep learning* model training. This study applied four separate image quality improvement techniques, namely *Histogram Equalization* (HE), *Contrast Limited Adaptive Histogram Equalization* (CLAHE), *Gamma Correction*, *Contrast Stretching*. Each image in the dataset was processed with the four techniques individually to evaluate the influence of each technique on classification performance.

1. Histogram Equalization (HE)

Histogram Equalization (HE) is a contrast enhancement technique that aims to spread the distribution of pixel intensity evenly across the available dynamic range. In HE, the transformation is performed based on the *Cumulative Distribution Function* of the image histogram. The general formula for HE is represented through equation (1).

$$S_k = (L-1)\sum_{j=0}^{k} p_r(r_j) \quad (1)$$

with is the value of the new intensity, L is the sum of the intensity level (e.g. 256 for 8-bit), and is the probability of the occurrence of the intensity $s_k p_r(r_j) r_j$

2. Gamma Correction

Gamma Correction is a technique used to adjust the brightness level of an image by changing the relationship between the input and output intensity values nonlinearly. This technique aims to optimize visual lighting to better match the characteristics of human perception of light. Mathematically, the gamma transform is expressed in equation (2).(Fitriyah & Wihandika, 2021)

$$I_{Output} = I_{Input}^{\gamma} \quad (2)$$

where and is the intensity of the pixels before and after the transformation, and γ is the gamma parameter (>1 for darkening, <1 for brightening). $I_{output} I_{input}$ To identify the optimal luminance mapping for feature extraction, an ablation study was conducted using gamma (ϒ) values of **0.5, 1.0, and 2.0**. A ϒ < 1.0 (0.5) was used to brighten underexposed images, ϒ > 1.0 (2.0) to darken overexposed areas, and ϒ = 1.0 served as a baseline control.

3. Contrast Stretching

Contrast Stretching is a linear method to expand the range of pixel intensity in an image to better meet the available dynamic range. This technique aims to improve contrast by clarifying the difference between light and dark pixels. The basic formula for (Lubis & Darmawan, 2021)*contrast stretching* is expressed in equation (3).

$$I' \frac{I - I_{min}}{I_{max} - I_{min}} \times (N_{max} - N_{min}) + N_{min} \quad (3)$$

with *I* being the initial intensity, *I'* being the resulting intensity, *Imin* and *Imax* being the minimum and maximum intensities in the image, and N_{min} and N_{max} being the new desired intensity range.

4. Adaptive Contrast Enhancement (CLAHE)

The increase in adaptive contrast in the image in this study was carried out using *the Contrast Limited Adaptive Histogram Equalization* (CLAHE) method. CLAHE is a development of the *Histogram Equalization* (HE) technique which aims to improve the quality of contrast in a more controlled manner, especially in images that have uneven lighting. The basic technique of HE works by calculating the *Cumulative Distribution Function* (CDF) of the pixel intensity histogram, which is represented in the form of an equation (4).

$$S_k = (L-1)\sum_{j=0}^{k} \frac{\mathcal{P}_r(r_j)}{N} \quad (4)$$

With *Sk* representing the value of the new pixel after the transformation process. In this formula, *L* indicates the number of intensity levels available, where for 8-bit images the value is 256. The *variable Pr*(*rj*) is the probability of the occurrence of the intensity value *rj* which reflects the frequency distribution of the pixel value in the image. While *N* expresses the total number of pixels present in the image(Buriboev et al., 2024). However, direct application of HE to the entire image often results in an *over-enhancement* effect, especially in areas that already have high contrast. To overcome this, CLAHE divides the image into several small parts or tiles, then equalizes *locally* on each of these parts. In addition, CLAHE applies clip limits to limit the maximum height of the histogram to avoid dominance of certain pixels. This *clipping* process is expressed by equation (5).

$$\beta = \frac{M}{N}\left(1 + \frac{\alpha}{100}(S_{max} - 1)\right) \quad (5)$$

Parameters *M* and *N* are variables used in ratio calculations, where α (also called *clipfactor*) is valued between 0 and 100 and acts as a scaling factor. *Smax* represents the maximum value of pixels in an image (Alzami et al., 2024). After the *equalization process*, to avoid striking differences between tiles, bilinear interpolation is performed so that the transitions between parts of the image are smooth and do not cause a block effect. In this study, the implementation of CLAHE was applied to the *luminance*

channel (L) of the LAB color space, considering that this channel represents the intensity of light in the image. Once the equalization process is performed, the image is returned to the RGB color space. Furthermore, to determine the best trade-off between contrast and noise, we evaluated clip limits of 2.0, 3.0, and 4.0. The clip limit parameter is critical in sky imagery to prevent the over-amplification of noise in homogeneous (flat) atmospheric regions.

5. Resizing & Preprocessing for MobileNetV3

Before the imagery can be rendered to the MobileNetV3 architecture, it requires a dimension adjustment and pixel value normalization stage to match the network input format and maintain consistency during the data training process. MobileNetV3, like other CNN architectures, requires a fixed-sized input of 224×224 pixels and 3 color channels (RGB). The *resizing process* is carried out using the interpolation function of *OpenCV* to adjust the size without causing major distortion to the visual content of the image.

After the resizing, the image then normalizes using *the preprocess_input()*function of the *tf.keras.applications.mobilenet_v3 module*. This function automatically converts the pixel value from the range [0, 255] to the range [-1, 1], in accordance with the preprocessing standards used during the initial training of MobileNetV3. This normalization assists the network in accelerating convergence and increasing stability during training. The transformation of the pixel value is carried out based on equation (6).

$$I' = \frac{I}{127.5} - 1 \tag{6}$$

Where I is the value of the original pixel intensity, which is on a scale of 0 to 255, where 0 represents black and 255 represents white. The *I' value* is the result of normalization that transforms the pixel value into an interval of -1 to 1. This normalization aims to improve the stability and convergence of training/prediction, as well as align the input distribution with the data distribution at the time the model is trained (Alzami et al., 2024). By applying four separate preprocessing techniques, this study aims to evaluate the contribution of each method to image quality and performance of the Air Quality Index (AQI) classification based on sky imagery.

Splitting Data

After all the images are completed through the preprocessing process, the next stage is to divide the dataset into two main subsets, namely the *training data* and the test *data*. This division aims to ensure that the evaluation of model performance is carried out objectively against data that has never been seen before during the training process, thereby avoiding *overfitting* and providing a realistic picture of the model's generalization capabilities. In this study, the split ratio used was 80:20, of which 80% of the data was used to train the model, and the remaining 20% was used for testing. The distribution process was carried out randomly but controlled so that the proportion of class distribution remained balanced between the training data and the test data (*stratified sampling*).

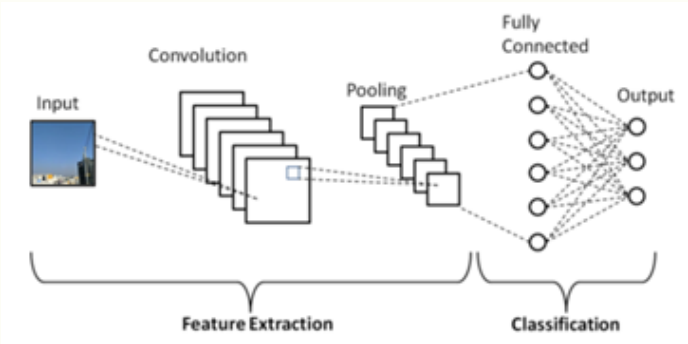
This is especially important in the case of image classification with uneven class distribution, so that the model is not biased against the majority class. *Data splitting* is done using the *scikit-learn library*, specifically the *train_test_split() function* which allows for data randomization and separation with customized parameters. With this method, the model gains the opportunity to learn from the majority of the data, while also being tested with data it has never known before, to measure its accuracy and generalization ability in a fair and measurable manner.

Model Development

1. Convolutional Neural Network

CNN is a *deep learning architecture for* grid-like, *image-like data processing, which extracts spatial features through convolutional layers with* non-linear activation functions (Azmi et al., 2023). Its advantage lies in its ability to extract features automatically and hierarchistically, from simple to complex features. The main components of CNN include *convolutional layers*, *pooling layers*, and *fully connected layers* that work sequentially to form abstract representations of images. The structure of this filter is arranged in such a way that it can form filters with dimensions of length and width (pixels). This filtering process

is carried out at various CNN layers during the learning stage as shown in Figure 3.(Zhao et al., 2024)



Picture 3. Convolutional Neural Network

2. Convolutional Layer

The convolutional layer on CNN extracts local features from the image by applying filters to detect visual patterns such as edges, textures, or shapes. The convolutional operation is described through equation (7).

$$S(i,j) = (X * K)(i,j) = \sum_m \sum_n X(i+m, j+n) \cdot K(m,n) \quad (7)$$

Where *X* is the input image, *K* is the kernel/filter, and *S(i,j)* is the convolution result at position *(i,j).* This process results in a more dense and informative representation of features, and by increasing the number of filters and convolutional layers, the model can learn more complex features.(Ajit et al., 2020)

3. Rectified Linear Unit (ReLu)

ReLU is a common activation function in CNN that overcomes *the vanishing gradient* by changing a negative value to zero and retaining a positive value. The function of ReLU is defined in equation (8).

$$f(x) = \max(0, x) \quad (8)$$

This function works by returning a value *of x* if *x* is positive ($x > 0$) and returning 0 if *x* is negative or zero ($x \leq 0$). The implementation of ReLU after convolution makes the model more non-linear and accelerates convergence during training. (Zhao et al., 2024)

4. Pooling Layer

Pooling layer reduces the spatial dimension of convolutional extraction features, reduces parameters and computation, and controls *overfitting*. A commonly used type is *Max Pooling*, which is represented in equation (9).

$$f(x) = \max_{i \in R} x_i \quad (9)$$

Where *R* is the local region of the map feature. Alternatively, *Average Pooling* calculates the average of the elements within a local area, which is formulated in equation (10).

$$f(x) = \frac{1}{|R|} \sum_{i \in R} x_i \quad (10)$$

Where *R* is the local region in the map feature and *xi* is the pixel value within that region. *Average pooling* has the advantage of keeping background information or areas with low contrast.

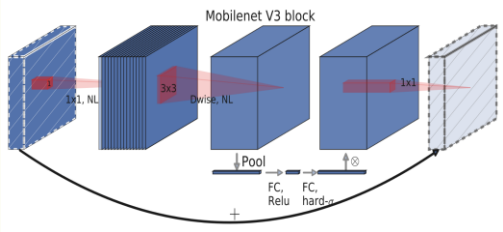
5. Pretrained Weights

Pretrained weights are weights that have been trained on large datasets and reused for new tasks, speeding up training and improving accuracy. The use of *pretrained weights* from *ImageNet* on MobileNetV3 in this study speeds up training and improves performance, especially on small datasets, compared to training from scratch.

6. MobilenetV3

MobileNetV3, a development of MobileNetV1 and V2, focuses on computing efficiency and high performance. It uses *depthwise separable convolution*, *Squeeze-and-Excitation* (SE) *block*, and *h-swish* as activation functions. There are two variants: MobileNetV3-Small for limited computing, and MobileNetV3-Large for higher accuracy. This study uses MobileNetV3-Large, which combines *inverted residual blocks* and SE blocks, suitable for image classification on a limited platform. Figure 4 shows the MobileNetV3 architecture that has been tuned for the classification of sky imagery.(Qian et al., 2021)

Baseline Model: VGG16 To validate the architectural efficiency and performance of MobileNetV3, this study employs VGG16 as a comparative baseline model. VGG16 is a classic Convolutional Neural Network architecture featuring 16 weight layers with a significantly larger parameter count (~138 million) compared to MobileNetV3. The inclusion of VGG16 aims to measure the trade-off between accuracy and computational efficiency, and to investigate whether the image enhancement techniques (such as HE) yield consistent positive impacts across different architectures.



Picture 4. MobilenetV3 Block

7. Hyperparameter Tuning / Fine Tuning

Hyperparameter Configuration and Fine-Tuning The training strategy was designed to maximize transfer learning efficiency. The MobileNetV3 architecture was modified by appending Fully Connected layers (1024 and 512 units) equipped with a Dropout rate of 0.3. This dropout value was selected to mitigate overfitting given the limited dataset size. For fine-tuning, we unfroze the last 50 layers of the base model. This approach allows the model to adapt high-level features specific to air pollution patterns, while retaining the robust low-level features learned from ImageNet.

Optimization was performed using the Adam optimizer with an initial learning rate of 0.0001. This conservative learning rate was deliberately chosen to preserve the stability of the pre-trained weights, preventing drastic updates that could degrade existing feature representations. A batch size of 16 was selected to balance memory efficiency on the NVIDIA T4 GPU with gradient convergence stability. Training was conducted for a maximum of 150 epochs utilizing an EarlyStopping mechanism (patience of 15) to automatically halt training if validation loss stagnated, thereby ensuring computational efficiency.

Model Training

The model training process is carried out using data train and tests that have gone through the pre-processing stage. The pre-prepared MobileNetV3Large model is trained with *predefined epoch* and *batch sizes*. To optimize training performance and avoid *overfitting*, three main *callbacks* are applied, namely *EarlyStopping*, *ModelCheckpoint*, and *ReduceLROnPlateau*. *EarlyStopping* is used to automatically stop the training process if the *validation loss* value does not show a decrease for 15 *consecutive epochs*. *ModelCheckpoint* stores the best model weights based on the lowest *validation loss* value achieved during the training process. Meanwhile, *ReduceLROnPlateau* is used to adjust *the learning rate* value adaptively, by reducing by 50% when *validation loss* is stagnant, thus helping the model to exit the local minima and accelerate convergence.

The training was conducted using *the model.fit() method*, with inputs in the form of pre-processed image data *(x_origin_train)* and the target value of the air quality index *(y_train).* The entire training history, including loss values and other metrics in the training and validation data, is stored in the history variable that will be further analyzed at the model evaluation stage.

Model Evaluation

1. Precision, Recall, F1-Score

Evaluation metrics such as *precision, recall, and F1-score* are used to evaluate the performance of the model's classification in identifying air quality based on sky images. *Precision* measures the proportion of correct positive predictions, while *recall calculates the* proportion of the total positive data that the model successfully recognizes. *F1-score* is a harmonization between *precision* and *recall*, which is especially useful when class distributions are unbalanced. The formula of each is in equation (11)(12)(13)(14).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

$$Precision = \frac{TP}{TP+FP} \quad (12)$$

$$Recall = \frac{TP}{TP+FN} \quad (13)$$

$$F1Score = 2x\frac{Precision\ x\ Recall}{Precision + Recall} \quad (14)$$

where TP is *true positive*, TN is *true negative,* FP *false positive*, and FN *false negative*. [23]

2. Confusion Matrix

Confusion matrix is a visual representation of the model's prediction of actual values, especially in classification tasks. This matrix presents the number of correct and false predictions for each class. The main components in the confusion matrix are *true positive* (TP), *false positive* (FP), *true negative* (TN), and *false negative* (FN), which are used in calculating metrics such as *precision* and *previous recall*. Confusion matrix visualization *helps* in identifying model biases against a particular class, as well as to understand the types of errors that are most frequently made.(Heydarian et al., 2022)

3. Grafik Loss & Accuracy

During the training process, the *loss* and *accuracy values* of the training and validation data were recorded in each *epoch*. The loss graph shows how well the model minimizes errors, while the accuracy graph represents how often the model makes correct predictions. The patterns of these two graphs are important indicators for detecting symptoms of *underfitting* or *overfitting*.

4. RMSE & R^2-score

To evaluate the regression performance of the model against *the Air Quality Index* (AQI) value, the *Root Mean Squared Error* (RMSE) and *Coefficient of Determination* (R^2) metrics were used. The RMSE measures the predicted average deviation from the actual value in the same unit as the target, with equation (14).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad (15)$$

The notation *n* expresses the number of data samples, *yi* is the ith actual observation value, while y^i is the model's prediction value for the ith sample While R^2 measures the proportion of target variance that the prediction model successfully explains, presented through equation (15).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2} \quad (16)$$

The notation y^- *represents the average of yi's observational value.* A low RMSE value and an R^2 close to 1 indicate good regression performance.(Chicco et al., 2021)

RESULTS AND DISCUSSION

This chapter presents and discusses the results of a comparative experiment of four image enhancement techniques—*Histogram Equalization* (HE), *Contrast Limited Adaptive Histogram Equalization* (CLAHE), *Gamma Correction*, and *Contrast Stretching* on the performance of the MobileNetV3 model in

estimating AQI values and air quality category classification from sky images. The analysis includes visual evaluations, histograms, training performance, as well as quantitative evaluation metrics obtained from code execution.

Visual Analysis and Histogram of Image Preprocessing Results

Preliminary analysis was performed to understand the visual impact and changes in pixel intensity distribution caused by each pre-processing technique on the sample image.

1. Comparison Visualization

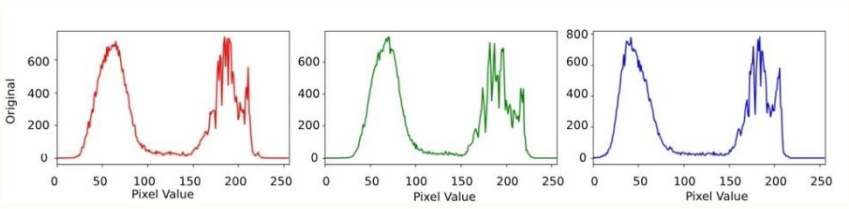
To provide a qualitative picture of the influence of each technique, a visual comparison was made between the original image and the pre-processed image. The results of a visual comparison of a sample image are presented in Figure 5

Picture 5. Comparison Visualization

Figure 5 shows an example of the original image and its transformation results using all four preprocessing techniques. Visual comparison images show that *Histogram Equalization* significantly improves global contrast, but can cause color artifacts or *noise* in some areas, as seen in bluish parts of vegetation. *Gamma Correction* (with *the default gamma*, i.e. 1.0) shows no drastic change from the original image in this sample. *Contrast Stretching* appears to improve the overall brightness and contrast, making the details of clouds and vegetation a little more vivid. CLAHE manages to improve local contrast, especially in cloud textures and details in darker areas, without changing color and excessive noise amplification as in HE.

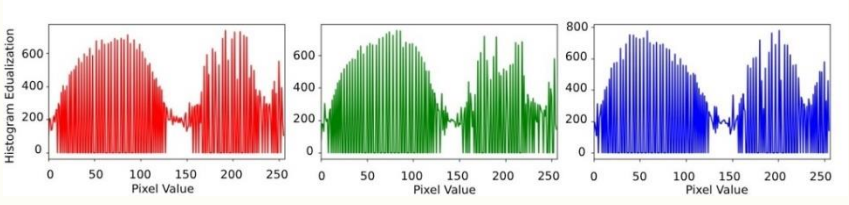
2. Histogram Analysis

For quantitative analysis, the pixel intensity histogram for each color channel (Red, Green, Blue) was calculated before and after the application of the image enhancement technique. Figures 6, 7, 8, 9, 10 present a comparison of these histograms.



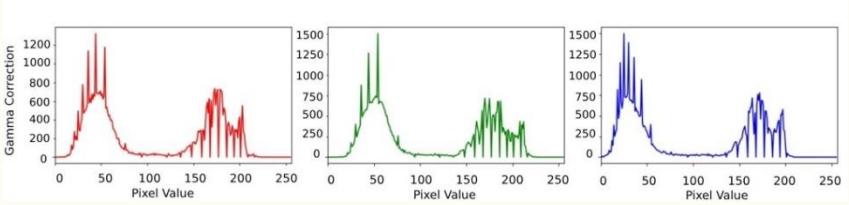
Picture 6. Histogram Analysis

The original image histograms for the R, G, and B channels seen in Figure 6 show a distribution of pixels that tend to be concentrated at a certain intensity range. Often, there are narrow peaks, indicating that most pixels have similar brightness values and that the dynamic range of the image has not been fully utilized.



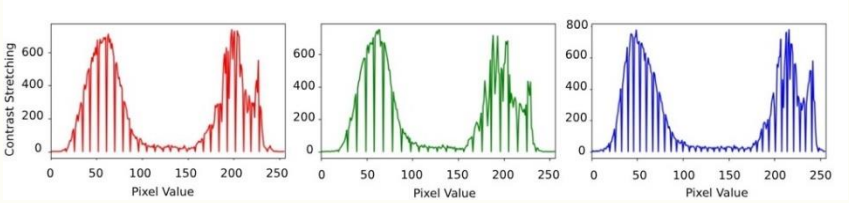
Picture 7. HE Histogram

The HE result histogram seen in Figure 7 shows a much more even distribution of pixels and is spread across the entire intensity range (0-255). This is the main characteristic of HE, which aims to "flatten" the histogram. Although increasing the global contrast, this forced spread can lead to the loss of information regarding the original intensity distribution and sometimes combine multiple previously separate peaks.



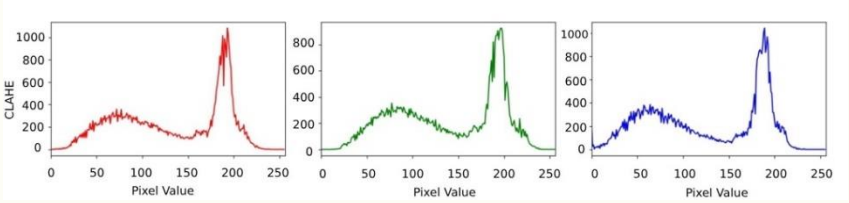
Picture 8. Histogram Gamma Corection

Changes in the histogram due to *Gamma Correction* such as in Figure 8 where the visuals appear darker will indicate a shift in pixel distribution. If the image is darkened (gamma > 1), the peaks of the histogram will shift to the left (towards a lower intensity value). On the other hand, if it is brightened, the crest will shift to the right.



Picture 9. Histogram Contrast Stretching

The histogram after Contrast Stretching seen in Figure 9 shows that the range of pixel values is expanded to include the entire spectrum from 0 to 255. The vertices on the original histogram are "stretched" but their relative shape and the position of their vertices tend to be retained. This stretch sharpens the difference between the minimum and maximum pixel values in the image.



Picture 10. Histogram CLAHE

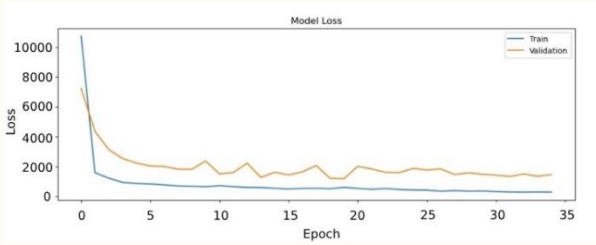
The CLAHE histogram as visualized in Figure 10 shows a wider distribution than the original image, indicating an increase in contrast. However, in contrast to HE, the CLAHE histogram does not become completely uniform. Because CLAHE works locally, it tends to increase contrast in areas where detail is needed while trying to maintain an overall intensity distribution so that it doesn't deviate too much from the original.

Training Results and Model Evaluation

1. Model Training Results

The MobileNetV3Large model with *transfer learning* is trained separately for each dataset of pre-processed results (Original, HE, Gamma, CS, CLAHE). A subset of data of 3000 samples was used for this comparative study. The training loss curve and validation provide insight into the model's learning process.

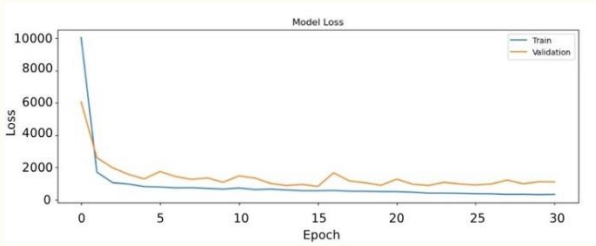
a. Model with Original Data



Picture 11. Model with Original Data

The *loss* curve shows a steady *decrease* in losses in both training and validation data, although *loss validation* shows greater fluctuations and is slightly above *training losses*, indicating a good learning model but there is a slight gap in generalization.

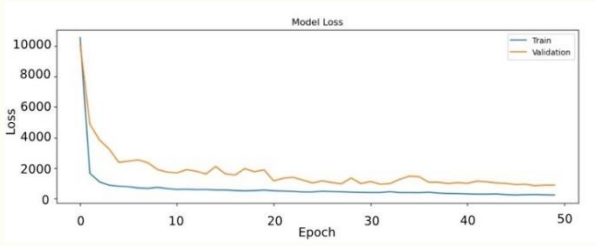
b. Model with Histogram Equalization (HE)



Picture 12. Model with Histogram Equalization (HE)

The *loss* curve also shows a good downturn, with loss validation relatively close to loss training and more stable than the original model after the initial few epochs. This suggests HE might help the model in generalization.

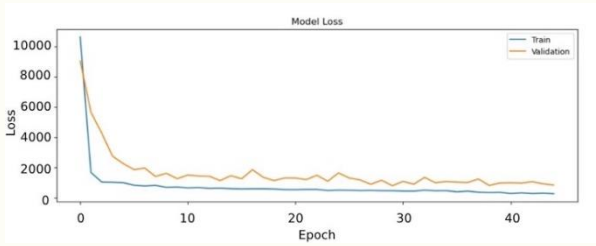
c. Models with Gamma Correction



Picture 13. Models with Gamma Correction

The loss curve shows a similar convergence pattern to the original model, with a rapid decline in losses at the beginning and stabilization in later *epochs*. The gap between *training loss* and *validation* appears to be consistent.

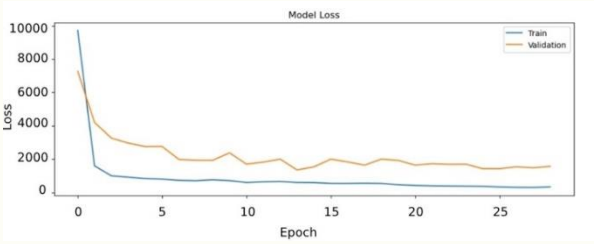
d. Model dengan Contrast Stretching



Picture 14. Model dengan Contrast Stretching

The loss curve shows a rapid decline in the early epochs, signaling an effective learning model. Although both curves continued to descend, the validation curve was consistently above the training curve and showed greater fluctuations, indicating a slight *overfitting*, although this technique was previously shown to provide the highest category classification accuracy.

e. Models with CLAHE



Picture 15. Models with CLAHE

This learning pattern (rapid decline, slight *overfitting*, stabilization) suggests that although CLAHE helps the model achieve good regression performance. Overall, however, the training process was successful, and the model showed good generalization skills.

2. Model Performance Evaluation

Quantitative evaluation was performed on the *test set* to compare the performance of the trained model with the data of different pre-processing results. The results of the comparison of the performance of classification metrics and regression metrics are summarized in Table 4.

Table 4. Model Performance Evaluation

Metric	Original	Histogram Eq. (HE)	CLAHE	Gamma Correct	Contrast Stretch (CS)
Classification					
Accuracy (%)	0.916	0.923	0.893	0.856	0.916
F1 Score	0.949	0.955	0.938	0.918	0.952
Precision	0.910	0.928	0.890	0.852	0.913
Recall	0.991	0.984	0.991	0.995	0.996
Category Accuracy	0.643	0.690	0.583	0.540	0.613
Regression					
MSE (m^2)	935.742	915.697	2016.843	1745.349	1766.691
RMSE	30.589	30.260	44.909	41.777	42.032
IT IS	20.401	20.594	27.862	30.574	28.977
R^2	0.904	0.913	0.852	0.820	0.826

1. Regression Performance (Estimated AQI value)

Based on the results of the evaluation, the Histogram Equalization (HE) technique shows the best regression performance consistently. HE managed to achieve the highest R^2 value of 0.913, which indicates the model's best ability to explain the variability of AQI data. This advantage is also supported by the lowest error score in the *Mean Squared Error* (MSE) metric of 915,697 and *the Root Mean Squared Error* (RMSE) of 30,260. In addition, the prediction versus actual plot for HE visualized in Figure 16 will show a tighter distribution of points around the diagonal line, confirming superior prediction accuracy.

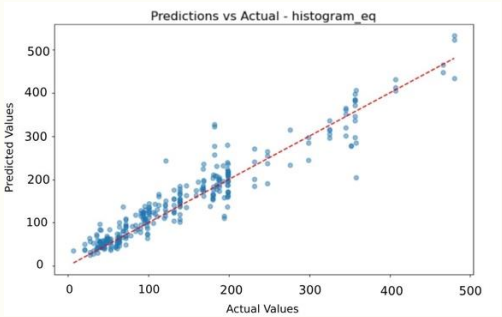


Figure 16. Regression Performance (*AQI Estimate*)

2. AQI Category Classification Performance

In the classification metric, Histogram Equalization (HE) once again demonstrated its superiority as the most effective pre-processing technique. This technique results in *the highest Category Accuracy* of 0.690, which means that HE is most accurate in classifying images into the correct AQI category (e.g., *Good*, *Moderate*, *Unhealthy*). This advantage is also strengthened by the acquisition of the highest scores in the Accuracy (0.923), F1 Score (0.955), and Precision (0.928) metrics. In addition, *the confusion matrix* for HE in Figure 17 will show a good balance in minimizing misclassification.

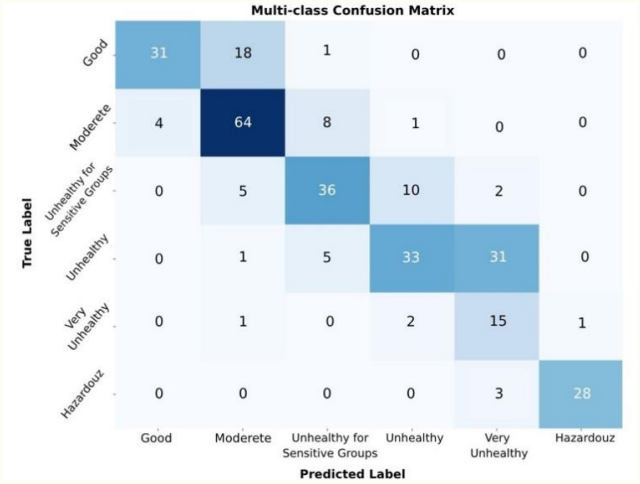


Figure 17. Regression Performance (AQI Estimate)

A deeper analysis of the Confusion Matrix (Figure 17) reveals that the model exhibits excellent sensitivity for extreme classes. The 'Good' and 'Hazardous' categories were identified with high precision, with minimal misclassification cases. The primary challenge is evident in the transition classes, particularly between the 'Unhealthy' and 'Very Unhealthy' categories, where prediction interchange occurred. This can be attributed to the high similarity in visual features (haze levels) between these two classes. Nevertheless, the HE technique proved capable of accentuating distinguishing features in these intermediate classes compared to raw imagery, thereby minimizing overall visual ambiguity.

Based on its dominant prominence on almost all regression and classification metrics, Histogram Equalization (HE) was selected as the best pre-processing method, and models trained on HE data were kept as the final model.

3. Comparative Analysis with Baseline Model

Table 5. Comparison of Best Performance between MobileNetV3 and VGG16

Model	Best Preprocessing	R2	MSE	Category Accuracy
MobileNetV3	Histogram Eq	0.9137	915.69	0.6900
VGG16	Histogram Eq	0.7011	2863.22	0.4267
VGG16	Contrast Stretching	0.8550	1745.07	0.4700

Table 5 presents a direct comparison between the proposed model (MobileNetV3) and the baseline model (VGG16). The results unexpectedly reveal that MobileNetV3 significantly outperforms VGG16 across all key metrics. While MobileNetV3 achieved its peak performance using Histogram Equalization (HE) with an R^2 of 0.913 and the lowest MSE of 915.69, VGG16 struggled to generalize, achieving a maximum R^2 of only 0.855 with Contrast Stretching. Notably, applying HE to VGG16 resulted in a performance drop (R^2 0.701), indicating that the lightweight architecture of MobileNetV3 is better suited for adapting to the enhanced features provided by global histogram equalization in this specific domain.

4. Error Analysis

To understand the limitations of the model, a case of misclassification was analyzed in which an image with the original category "Unhealthy" was incorrectly predicted as "Very Unhealthy" with an AQI value of 201.52 displayed in Figure 18.

Figure 18. Examples of Model Classification Errors

This error occurs due to visual ambiguity in the category boundary. The physical characteristics of the sky such as the thickness of the haze for the upper level of "Unhealthy" and "Very Unhealthy" conditions of the lower level can be very similar, making it difficult for the model to distinguish. Although it is categorically incorrect, it should be noted that the predicted value of AQI 201.52 is numerically not far from the threshold (~200). This shows that the model still manages to recognize high levels of pollution, but misses only slightly in precise category placement.

Discussion/Comparisan and Discussion

The results of the experiment clearly show that the selection of image enhancement techniques has a significant impact on the performance of the MobileNetV3 model.

1. Dominance of Histogram Equalization (HE): Recent results show that Histogram Equalization (HE) is the most effective technique comprehensively, excelling in almost all regression and classification metrics. The strong global contrast enhancement by HE, evidenced by the even distribution of the histogram, seems to have succeeded in highlighting crucial visual features in sky imagery that correlate with pollution levels. This helps the model not only in estimating precise AQI values (highest R^2 0.913) but also in accurately distinguishing boundaries between categories (highest Category Accuracy 0.690).
2. Performance Consistency for Regression and Classification: Interestingly, these recent results show that the optimal technique for regression (HE) turns out to be the optimal technique for classification. This implies that the visual features enhanced by HE—such as cloud texture, turbidity level, and sky color gradation—are fundamental and equally relevant to both tasks, both for predicting numerical values and determining their categories.
3. Importance of Pre-processing: The advantages of HE become more pronounced when compared to the performance of models that use native data. There was significant improvement in key metrics such as R^2 (from 0.904 to 0.913) and Category Accuracy (from 0.643 to 0.690). This strongly confirms

that the proper implementation of pre-processing is not only beneficial, but crucial to maximize the potential of deep learning models in this task.

Implications and Contributions: This study confirms the hypothesis that image enhancement techniques can improve model accuracy. The main contribution of this study is the identification of Histogram Equalization (HE) as a superior and consistent pre-processing technique for image-based air quality estimation tasks using MobileNetV3. These findings provide a more straightforward practical guide: HE is the primary choice for achieving optimal performance in both AQI value prediction and category classification. Limitations related to the dataset remain, but the justification for the similarity of atmospheric conditions supports the relevance of this finding. The validity of the Histogram Equalization (HE) technique's superiority is further confirmed through result consistency across all evaluation metrics. HE not only excelled in a single aspect but statistically dominated both regression metrics (R^2, RMSE) and classification metrics (Accuracy, F1-Score). This consistent performance improvement across various measurement parameters indicates that the enhancement offered by HE is significant and robust against data variations, rather than being a mere result of random variance from the stochastic training process.

CONCLUSION

This study aimed to evaluate the influence of various image enhancement techniques on the performance of the MobileNetV3 model in predicting air quality and found that the Histogram Equalization (HE) technique consistently provided the best performance. This comprehensive advantage is evident in the regression task with the highest R^2 value (0.913), as well as in the classification task with the highest Category Accuracy (0.690). These results significantly show that for this case, there is no trade-off between regression accuracy and classification, but rather that one technique excels in both evaluation domains.

Based on these findings, the study provides straightforward practical recommendations: Histogram Equalization (HE) is the most effective pre-processing technique to maximize the accuracy of the MobileNetV3 model in image-based air quality estimation tasks. This success reinforces the crucial role of the pre-processing stage in the deep learning pipeline and confirms that the combination of classic image-enhancing techniques with efficient modern architecture can be a highly effective approach for future air quality monitoring solutions. To ensure research sustainability, future work should focus on training the model with larger and more heterogeneous datasets from diverse geographical locations to enhance generalization capabilities. Furthermore, the development of a mobile or web-based prototype is recommended as a proof of concept for real-time air quality monitoring in real-world environments.

REFERENCES

Ajit, A., Acharya, K., & Samanta, A. (2020). A Review Of Convolutional Neural Networks. *IEEE Xplore,* 1–5. https://doi.org/10.1109/ic-ETITE47903.2020.049

Alzami, F., Naufal, M., Al Azies, H., Winarno, S., & Soeleman, M. A. (2024). Time Distributed MobileNetV2 with Auto-CLAHE for Eye Region Drowsiness Detection in Low Light Conditions. (*IJACSA) International Journal of Advanced Computer Science and Applications*, 15(11), 488-500. https://dx.doi.org/10.14569/IJACSA.2024.0151146

Amahoru, A. H., Dulhasyim, A. B. P., & Pulu, S. R. (2024). Analisis Citra Visual Fase-Fase Bulan dalam Tinjauan Sistem Koordinat Bola Langit. *Jurnal Pendidikan Mipa*, *14*(1), 114–123. https://doi.org/10.37630/jpm.v14i1.1492

Azlansyah, M., Setiyono, B., & Komputasi, F. M. (2019). Penyisipan Pesan pada Citra Digital Menggunakan Metode Least Significant Bit. *Jurnal Sains Dan Seni ITS*, *8*(1), 2337-3520.

Azmi, K., Defit, S., & Sumijan, S. (2023). Implementasi Convolutional Neural Network (CNN) Untuk Klasifikasi Batik Tanah Liat Sumatera Barat. *Jurnal Unitek*, *16*(1), 28–40. https://doi.org/10.52072/unitek.v16i1.504

Bismi, W., & Qomaruddin, M. (2023). Klasifikasi Citra Genus panthera Menggunakan Pendekatan Deep learning Berbasis Convolutional Neural network (CNN). *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *5*(2), 172–179.

Buriboev, A. S., Khashimov, A., Abduvaitov, A., & Jeon, H. S. (2024). CNN-Based Kidney Segmentation Using a Modified CLAHE Algorithm. *Sensors*, *24*(23), 1-24. https://doi.org/10.3390/s24237703

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Coefficient Of Determination R-Squared Is More Informative Than SMAPE, MAE, MAPE, MSE And RMSE In Regression Analysis Evaluation. *Peerj Computer Science*, *7*, 1-24. https://doi.org/10.7717/peerj-cs.623

Dede, M., Widiawaty, M. A., Nurhanifah, N., Ismail, A., Artati, A. R. P., Ati, A., & Ramadhan, Y. R. (2020). Estimasi Perubahan Kualitas Udara Berbasis Citra Satelit Penginderaan Jauh Di Sekitar PLTU Cirebon. *Jambura Geoscience Review*, *2*(2), 78–87. https://doi.org/10.34312/jgeosrev.v2i2.5951

Enggari, S., Sumijan, & Tajuddin, M. (2024). The Development Of Color Histogram Method To Identify Air Quality Index Based On Sky Images. *Indonesian Journal of Electrical Engineering and Computer Science*, *34*(1), 186–196. https://doi.org/10.11591/ijeecs.v34.i1.pp186-196

Fitriyah, H., & Wihandika, R. C. (2021). *Dasar-Dasar Pengolahan Citra Digital*. Universitas Brawijaya Press.

Heydarian, M., Doyl, T. E., & Samavi, R. (2022). MLCM_Multi-Label_Confusion_Matrix. *IEEE Access, 10,* 19083-19095. https://doi.org/10.1109/ACCESS.2022.3151048

Jochsen, E., & Handhayani, T. (2024). Pengenalan Bangunan Bersejarah Pura Dengan Arsitektur InceptionV3 dan Xception. *Jurnal Eksplora Informatika*, *14*(1), 1–11. https://doi.org/10.30864/eksplora.v14i1.1064

Kamil, G. Z. (2022). Dilated Convolutional Vision Transformer dalam Kasus Rekonstruksi Citra Image Companding untuk Peningkatan Kualitas Citra Digital.

Lubis, L., & Darmawan, D. (2021). Perbandingan Metode Contrast Stretching. http://repository.uinsu.ac.id/id/eprint/14936

Mukundan, A., Huang, C. C., Men, T. C., Lin, F. C., & Wang, H. C. (2022). Air Pollution Detection Using a Novel Snap-Shot Hyperspectral Imaging Technique. *Sensors*, *22*(16), 6231. https://doi.org/10.3390/s22166231

Qian, S., Ning, C., & Hu, Y. (2021). MobileNetV3 for image classification. *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, 2021,* 490-497. https://doi.org/10.1109/ICBAIE52039.2021.9389905

Rokom, R. (2023). *Polusi Udara Sebabkan Angka Penyakit Respirasi Tinggi*. Sehat Negeriku.

Rouniyar, A. (2024). *Air Pollution Image Dataset from India and Nepal*. Kaggle.

Samsami, M. M., Shojaee, N., Savar, S., & Yazdi, M. (2019). Classification Of The Air Quality Level Based On Analysis Of The Sky Images. *2019 27th Iranian Conference on Electrical Engineering (ICEE)*, 1492–1497. https://doi.org/10.1109/IranianCEE.2019.8786738

Septiyana, D., Sukmono, A., & Yusuf, M. A. (2023). Pemantauan Kualitas Udara ISPU (PM10, SO2, NO2) Menggunakan Citra Landsat 8 dan 9 untuk Kecamatan Mijen Selama Pandemi Covid-19. *Jurnal Geodesi Undip*, *12*(3), 271–280.

Utomo, S., Rouniyar, A., Jiang, G. H., Chang, C. H., Tang, K. C., Hsu, H. C., & Hsiung, P. A. (2023). Eff-AQI: An Efficient CNN-Based Model for Air Pollution Estimation: A Study Case in India. *ACM International Conference Proceeding Series*, 165–172. https://doi.org/10.1145/3582515.3609531

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A Review Of Convolutional Neural Networks In Computer Vision. *Artificial Intelligence Review*, *57*(4). https://doi.org/10.1007/s10462-024-10721-6