AI Estimation of Air Pollution Percentage in Kyrgyzstan

Parvani Mokhammad * and Mohd Tauheed Khan

Posted Date: 29 April 2026

doi: 10.20944/preprints202604.2086.v1

Keywords: air pollution; Bishkek; convolutional neural network; deep learning; PM2.5; transfer learning; urban images

 doi:10.20944/preprints202604.2086.v1

Article

AI Estimation of Air Pollution Percentage in Kyrgyzstan

Mokhammad Parvani Vafa * and Mohd Tauheed Khan

Department of Computer Science, Ala-Too International University (AIU), Bishkek, Kyrgyzstan
* Correspondence: parvanivafa.mokhammad@alatoo.edu.kg

Abstract

Air pollution poses a serious environmental and public health problem in Bishkek, Kyrgyzstan, especially during the winter months when the concentration of particulate matter increases dramatically. Despite the urgency of the problem, there are fewer than eight monitoring stations in the city, which leaves large urban areas without proper air quality control. This article presents the first systematic study of image-based AQI assessment for Bishkek, which explores whether transfer learning models can extract visual cues related to environmental pollution from on-site urban photographs under real-world uncontrolled conditions. Two hybrid deep learning architectures, VGG16 and EfficientNetB0, each augmented with scalar PM2.5 input data, were trained and evaluated on a locally collected dataset of 1,014 image pairs–AQI. EfficientNetB0 consistently outperformed VGG16 on all three evaluation indicators, reducing RMSE by 15.5% (66.49 vs. 78.71) and MAE by 16.6% (49.00 vs. 58.78). Both models demonstrated a partial predictive signal in the AQI range from low to moderate, confirming that visual features related to the atmosphere can be detected even based on small datasets from local sources. The performance limitations reflect the scale of the dataset and sparse sensor infrastructure, rather than the lack of a studied structure, which is consistent with similar pilot studies conducted under similar data constraints. This work establishes a basic and methodological framework for future image-based air quality monitoring in Central Asia and identifies key bottlenecks — the size of the dataset, tag interference caused by geographic mismatches in sensor images, and the density of monitoring stations - that should be addressed in future work.

Keywords: air pollution; Bishkek; convolutional neural network; deep learning; PM2.5; transfer learning; urban images

1. Introduction

Air pollution is considered a global problem for the environment and public health, especially in our time. According to the World Health Organization for Air Quality [1], prolonged exposure to fine particulate matter ($PM_{2.5}$) significantly increases the risk of cardiovascular and pulmonary diseases. Based on statistics from the Lancet Commission on Environmental Pollution and Health [2], premature mortality in young people and children associated with environmental pollution exceeds nine million deaths per year worldwide, making it one of the leading factors in the global burden of disease.

In Central Asia, and in particular in Kyrgyz Republic, urban air quality has noticeably deteriorated over the past few years. In Bishkek, the problem of air pollution has a pronounced seasonal character, due to a combination of coal burning for heating residential premises, vehicle emissions and temperature fluctuations. In winter, the concentration of pollutants often exceeds the permissible levels. Despite the severity of the situation, there are fewer than eight air quality monitoring stations in the city, some of which are supported by a private initiative rather than a centralized public infrastructure [3]. Such a number of stations does not allow for a full-fledged analysis at the district level and leaves a significant part of the urban area without adequate coverage.

 doi:10.20944/preprints202604.2086.v1

World Bank [4] has emphasized the macroeconomic aspect of the problem several times: an increased concentration of fine particulate matter leads to a noticeable increase in health care costs and a decrease in labor productivity. Given these complex environmental, epidemiological, and economic factors, there is a clear challenge to further explore scalable and low-cost complementary monitoring approaches that can expand the network of fixed stations.

Recent advances in deep learning and computer vision have demonstrated a potential path to this. Convolutional neural networks (CNNs), pre-trained on large arrays of natural images, can record visually monitored features of the atmosphere—haze, reduced contrast, decreased horizon visibility, and changes in color temperature—in the form of three-dimensional images that can be correlated with the concentration of particles in the air. Transfer learning [5] further facilitates the application of such methods in data-deficient environments by using concepts developed in ImageNet and adapting them to domain-specific regression goals.

This study investigates whether transfer learning models can extract pollution-related signals from the cityscapes collected in Bishkek. Two hybrid deep learning architectures based on VGG16 and EfficientNetB0 are compared using locally collected image data in conjunction with AQI measurements. The main results of this document are: (1) creation of a local air pollution dataset based on images for Bishkek; (2) evaluation of two regression models for transfer learning; (3) Analysis of constraints characteristic of an urban environment with rare observation; and (4) Creation of a framework for future research in Central Asia.

2. Related Work

2.1. Health Burden and Monitoring Urgency

There are numerous studies confirming the link between the concentration of $PM_{2.5}$ particles and high mortality, as well as development of chronic diseases. According to the GBD 2021 Risk Factors Collaborators [6] study, exposure to ambient PM is among the leading modifiable risk factors for cardiovascular and chronic respiratory diseases worldwide. Maji et al. [7] conducted a study that quantified the burden of disease caused by exposure to ambient $PM_{2.5}$ and PM_{10} in 130 Chinese cities and found that particulate matter exposure is one of the causes of premature death in humans, according for tens of thousands to hundreds of thousands of deaths per year, depending on the city's population and pollution intensity. These results confirm the importance of air quality monitoring, especially in cities like Bishkek, where monitoring is insufficient for accurate assessment.

2.2. Low-Cost Sensor Networks

Lower-cost electrochemical and optical particle sensors have been proposed as scalable complements to the stations. During their experiment, Kumar et al. [8] identified some calibration deviations, cross-sensitivity to humidity and temperature, and variability between departments as constant obstacles by analyzing the state of more budget-friendly sensors for urban air quality management. Liu et al. [9] implemented a two-channel CNN system on a low-cost on Raspberry Pi-based sensor platform for simultaneous classification and ambient particulate $matter_{2.5}$, the experiment showed improved performance that demonstrated the possibility of using advanced deep training technologies to monitor air quality, but despite their promising characteristics, inexpensive sensors still require physical deployment and maintenance, which limits their usefulness in resource-limited environments.

2.3. Deep Learning for Air Quality Prediction

Numerical forecasting using agrometeorological data, traffic data, or sensor data as input has been extensively studied using deep learning. Li et al. [10] applied ensemble tree-based methods and deep neural networks to estimate $PM_{2.5}$ based on forecasts from multiple sources across China, achieving R2 values exceeding 0.85 in well-censored areas. Zhao et al. [11] demonstrated that data fusion, which combines image features with numerical metadata for further training, provides a significant improvement in accuracy compared to single-modality approaches. That is, for clear

 doi:10.20944/preprints202604.2086.v1

numerical accuracy, a single data source for machine training is not enough. Gulia et al. [12] examined some urban air quality management strategies, highlighting the gap between the spatial detail required for effective analysis and that realistically provided by stationary monitoring networks.

Machine learning regression methods have also shown strong predictive performance for locally collected datasets in similar resource-constrained contexts.Burmachach et al. [13] applied Random Forest, CatBoost, and support vector machine models to estimate resale prices for cars from locally collected data in Kyrgyzstan and showed that ensemble and kernel-based regressors can effectively generalize even with limited and heterogeneous training data — a challenge directly analogous to the AQI regression task covered in the present study.

In the local academic context, computer vision techniques were applied at Ala-Too University to perform real-time detection tasks using the MediaPipe and OpenCV pipelines Esenalieva et al. [14], which confirms the possibility of using lightweight deep learning systems under limited resources. The present study extends this direction of environmental monitoring by applying transfer learning techniques to urban images to evaluate the AQI.

2.4. Image-Based AQI Estimation: Prior Art and Comparative Analysis

The literature, although small, contains a growing number of studies attempting to estimate pollution levels directly from visual images. Table 1 provides a brief description of the most comparable previous studies.

Table 1. Comparison of image-based and hybrid air quality estimation studies.

Study	Architecture	Images	Target	Best metric
[15]	CNN–LSTM (VGG/ResNet)	3,549	AQI	$R^2 = 0.94$
[16]	ResNet CNN	8,000+	AQI	$R^2 = 0.71$
[11]	Multimodal CNN	6,000+	AQI	$R^2 = 0.65$
[17]	Deep CNN + attention	15,000+	AQI	$R^2 = 0.78$
[18]	EfficientNet + metadata	12,000+	$PM_{2.5}$	MAE = 18 $\mu g/m^3$
[19]	CNN–LSTM (VGG16)	7,213	AQI	$R^2 = 0.92$
[20]	DCNN (smartphone)	∼1,000	PM2.5	—
[21]	AQI-Net (CNN+Grad-CAM)	11,000+	AQI	Acc = 99.8%
Present study	**VGG16 / EffNetB0**	**1,014**	**AQI**	**RMSE = 66.49**

Kow et al. [15] proposed a hybrid CNN–LSTM model for outdoor air images, achieving $R^2 = 0.94$ and RMSE = 5.38 for AQI estimation using 3,549 hourly labelled samples from a fixed monitoring station in Taiwan. Combining VGG and ResNet backbones with HSV colour statistics demonstrated that hand-crafted atmospheric colour features can complement deep visual representations. Zhang et al. [16] achieved $R^2 = 0.71$ using a ResNet-based CNN on over 8,000 outdoor images from multiple urban areas, applying explicit temporal stratification by time of day to reduce the influence of lighting conditions. Dong et al. [18] combined visual features extracted by EfficientNet with meteorological metadata (humidity, temperature, wind speed), achieving MAE = 18 $\mu g/m^3$ on a corpus of over 12,000 samples and demonstrating that auxiliary numerical inputs substantially reduce prediction error. Zhao et al. [11] reported $R^2 = 0.65$ using multimodal CNN fusion on over 6,000 labelled pairs. Xue et al. [17] used a deep CNN with spatial attention mechanisms on more than 15,000 images, achieving $R^2 = 0.94$.

Mondal et al. [20] estimated $PM_{2.5}$ and AQI from smartphone photographs collected in Dhaka, Bangladesh — a city characterised by severe seasonal pollution and a sparse monitoring infrastructure directly comparable to that of Bishkek. Using a deep convolutional neural network (DCNN) trained on approximately 1,000 outdoor images with locally assigned labels, the authors demonstrated that meaningful pollution signals can be extracted even from small, locally collected datasets. However, predictive accuracy remained limited, primarily due to label noise arising from the spatial mismatch between camera locations and the nearest monitoring stations. This finding is directly relevant to the present study: both the data-scale constraints and the label-noise mechanism identified by Mondal

 doi:10.20944/preprints202604.2086.v1

et al. [20] are structural features of our Bishkek dataset, providing independent evidence that the performance limitations we observe reflect the monitoring infrastructure rather than a failure of the modelling approach.

Wang et al. [19] proposed a hybrid CNN–LSTM architecture applied to surveillance-camera image sequences, achieving $R^2 > 0.92$ and RMSE < 8.5 for AQI estimation. Utomo et al. [21] proposed AQI-Net, a CNN-based model trained on more than 11,000 images from three Indonesian cities, augmented with Grad-CAM visualisation to identify which image regions most influence each prediction. The model achieved classification accuracy of 99.81%, illustrating that discretising the continuous AQI range into ordered categories can yield substantially higher task performance than direct regression, particularly when large training sets are available.

A clear pattern emerges: effective image-based AQI regression typically requires 5,000 to over 15,000 annotated image–label pairs, controlled or temporally stratified data collection, and often auxiliary numerical sensor inputs. The present study uses approximately one-fifth to one-fifteenth of this data volume under fully uncontrolled real-world conditions in a city previously unexplored in the literature. Despite this constraint, EfficientNetB0 reduces RMSE by 15.5% and MAE by 16.6% relative to VGG16, representing a consistent and reproducible internal improvement across all three evaluation metrics.

2.5. Research Gap

Existing image-based studies to assess air pollution mainly focus on data-rich regions with denser monitoring infrastructure and large publicly accessible datasets. Cities like Bishkek are still poorly understood despite heavy seasonal pollution and limited sensor coverage. This study closes this gap by presenting the first systematic assessment of image regression for environmental pollution based on the transfer learning method using locally collected city views of Bishkek.

3. Methodology

The dataset was collected manually by the primary author using an Apple iPhone 14 Pro Max held by hand, without a fixed mount or standardised camera angle. Images were captured across multiple districts of Bishkek as well as from a single reference location, covering the period from October 2024 to January 2025. All photographs were taken during daylight hours to ensure sufficient scene visibility. Concurrent AQI values were obtained from the IQAir platform (https://www.iqair.com), which reports real-time readings from the nearest available monitoring station to the photographer's location. A maximum temporal offset of 60 minutes was permitted between the timestamp of each photograph and the corresponding AQI reading. No meteorological metadata beyond the scalar $PM_{2.5}$ value was recorded at the time of collection. A total of 1,014 image–AQI pairs were obtained in this way, with each sample consisting of a geographically referenced photograph and the corresponding AQI value.

The dataset was deterministically divided into three subsets: 710 samples for training (70.0%), 151 for validation (14.9%), and 153 for testing (15.1%). All splits were performed at the image level to prevent data leakage.

The Table 2 contains partitioning statistics.

Table 2. Dataset partition statistics.

Split	Samples	Proportion	Usage
Training	710	70.0%	Model optimisation
Validation	151	14.9%	Early stopping, hyperparameter tuning
Test	153	15.1%	Final held-out evaluation
Total	**1,014**	**100%**	—

 doi:10.20944/preprints202604.2086.v1

3.1. Preprocessing

The size of all images has been changed to 224×224 pixels to achieve the expected input resolution of both backbone. For the values in VGG16 pixels, they were normalized by subtracting the average value using ImageNet, as implemented in `tf.keras.applications.vgg16.preprocess_input`. For EfficientNetB0, the built-in preprocessing function scales the values to $[0, 1]$ and applies additional internal normalization.

AQI target values were standardised using training-set statistics—mean μ_{AQI} and standard deviation σ_{AQI}—to obtain labels with zero center and unit variance, ensuring stable and fast convergence during training. The scalar $PM_{2.5}$ value was passed to the model in its original form as auxiliary input data. All predictions were de-standardised before computing the reported metrics, so the results are presented in the original AQI scale. A fixed initial random value (42) has been applied globally to TensorFlow, Numpy and the Random module in Python to ensure full reproducibility.

3.2. Model Architectures

3.2.1. VGG16 Hybrid Model

The VGG backbone [22] system, which has been pretrained in ImageNet with a remote classification header, processes the image input worth $224 \times 224 \times 3$ and creates a feature map worth $7 \times 7 \times 512$. The global mean pool layer (GAP) reduces this value to a 512-dimensional vector. The scalar input $PM_{2.5}$ is combined with this vector, resulting in a 513-dimensional collaborative representation. It passes through a dense layer of 128 units (ReLU activation) followed by a linear output neuron. The forward pass by the formula:

$$\hat{y} = W_{\text{out}} \cdot \text{ReLU}(W_{128} \cdot [\text{GAP}(\phi_{\text{VGG}}(I)) \parallel p]) \tag{1}$$

where I is the image, p is the $PM_{2.5}$ scalar, ϕ_{VGG} is the VGG16 feature extractor, and $\parallel$ denotes concatenation.

3.2.2. EfficientNetB0 Hybrid Model

An identical architecture was created using EfficientNetB0 [5] as the backbone. EfficientNetB0 creates a 1280–dimensional GAP. The concatenation and regression head $PM_{2.5}$ has the same design as the VGG16 variant.

3.3. Training Protocol

The training was conducted in two stages for each architecture. In **Phase 1** (feature extraction), all the base weights were frozen, and only the regression head was trained. The Adam optimizer was used with an initial training rate $\eta = 10^{-4}$, a Huber loss function, and a maximum of 15 epochs. Two callbacks were active: an early stop (waiting time = 5, tracking validation loss, restoring optimal weights) and a reduction in waiting time (coefficient = 0.5, waiting time = 3). Huber loss was preferred over MSE due to its resistance to AQI outliers.

In **Phase 2** (selective fine-tuning), the top four layers of each backbone were built up and the complete model was retrained with reduced training rate $\eta = 10^{-5}$ for up to 8 epochs with early stopping (patience = 3). This two–phase strategy prevents premature overwriting of pre-trained low-level features and at the same time enables task-specific adaptation in the upper layers. All experiments used a batch size of 32 and were run on an NVIDIA T4 GPU in Google Colaboratory.

3.4. Evaluation Metrics

Three regression metrics were computed on the test set after de-standardization:

 doi:10.20944/preprints202604.2086.v1

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{2}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

A negative R^2 indicates that the model performs worse than a constant mean predictor; this diagnostic value quantifies the gap that needs to be closed in future work.

Figure 1 illustrates the end-to-end pipeline of the proposed approach, from raw data collection through preprocessing, two-phase model training, and final evaluation on the held-out test set.

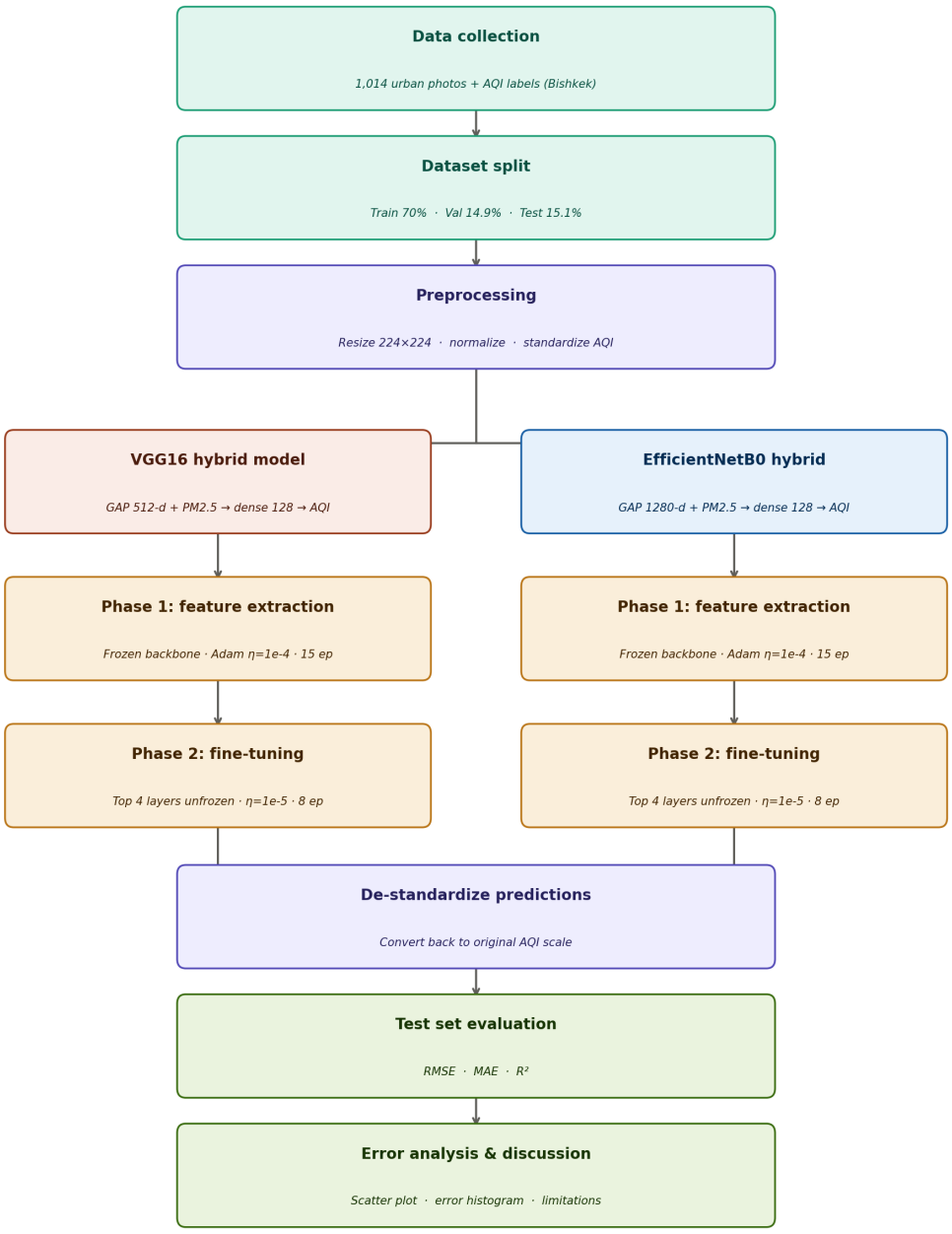

Figure 1. Methodology pipeline: from data collection to evaluation.

 doi:10.20944/preprints202604.2086.v1

3.5. Use of Artificial Intelligence Tools

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) for grammar checking and language editing of selected passages. All AI-generated suggestions were reviewed and edited by the authors. ChatGPT is not listed as an author. The authors take full responsibility for the final text.

4. Results

4.1. Training Dynamics

Figure 2 shows the Huber loss curves for VGG16 over 15 training periods (Phase 1). Training losses (blue curve) decrease monotonically from 1.61 in epoch 1 to 0.25 in epoch 15, reflecting the continuous adjustment of the head. Validation losses (orange curve) decrease from 1.81 in epoch 1 but remain at the level of 1.22–1.24 from about epoch 11, creating a constant and increasing gap between training and validation results. This pattern is a typical manifestation of overfitting caused by insufficient data volume compared to the capacity of the model.

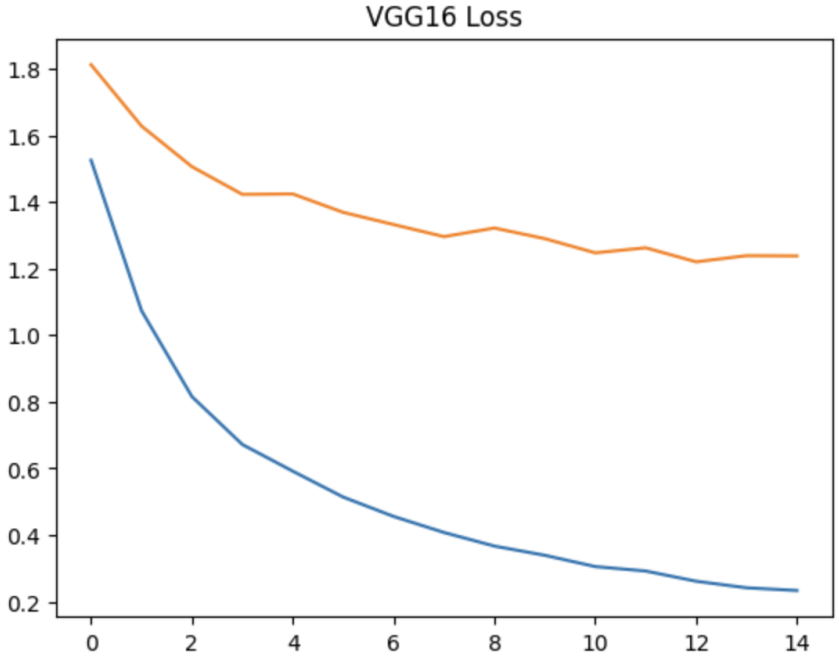

Figure 2. VGG16 Huber loss curves during Phase 1 training (15 epochs). Blue: training loss; orange: validation loss. The persistent gap between curves from epoch 3 onward indicates overfitting attributable to the limited dataset size.

Figure 3 shows the corresponding RMSE curves for VGG16.The training RMSE (normalized scale) decreases from 2.75 in epoch 1 to about 0.76 in epoch 15, while the default value of validation stabilizes by 2.06-2.07 after epoch 11. The final gap between the RMSE of training ($\approx$ 2.06) is 2.7, which confirms that the model effectively stored the training data instead of combining them into unseen samples.

 doi:10.20944/preprints202604.2086.v1

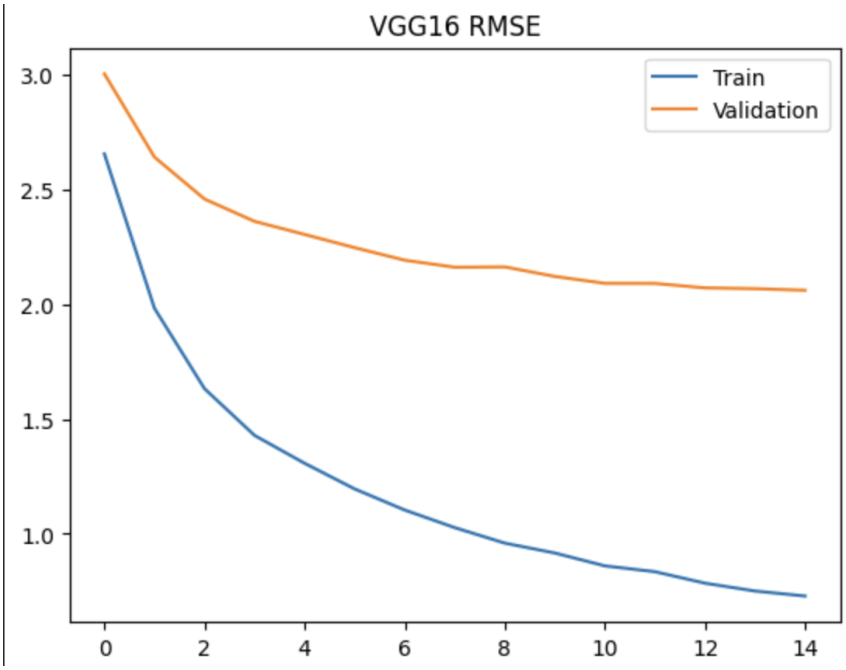

Figure 3. VGG16 RMSE curves during Phase 1 training. Blue: training RMSE; orange: validation RMSE. The gap widens continuously after epoch 3, corroborating the overfitting signal observed in the loss curves.

EfficientNetB0 showed a different convergence scheme. The initial validation loss (epoch 1:0.69) was significantly lower than that of VGG16 (epoch 1:1.81), reflecting more efficient reuse of features from pre-training. Validation loss for EfficientNetB0 stabilized around 0.67–0.69 by epoch 2 and showed minimal further improvement, suggesting that the backbone features quickly reached saturation. Table 3 compares representative per-epoch metrics for both models.

Table 3. Per-epoch training and validation metrics for both models during Phase 1 (selected epochs; normalised AQI scale).

	VGG16		EfficientNetB0	
Epoch	**Val Loss**	**Val MAE**	**Val Loss**	**Val MAE**
1	1.8117	2.2562	0.6921	1.1066
3	1.5057	1.9558	0.7002	1.1084
5	1.4238	1.8677	0.6903	1.0896
10	1.2896	1.7282	—	—
13	1.2200	1.6582	—	—
15	1.2378	1.6823	—	—

4.2. Test Set Evaluation

EfficientNetB0 achieved lower RMSE and MAE than VGG16 on the held-out test set. However, both models gave negative values of R^2, which indicates that the prognostic efficiency remained below the average for the current data set.

Table 4 reports the primary results on the held-out test set after de-standardization to the original AQI scale.

Table 4. Test set regression metrics for both models (AQI scale, de-standardised).

Model	**RMSE**	**MAE**	**R^2**
VGG16 + $PM_{2.5}$ hybrid	78.71	58.78	−0.794
EfficientNetB0 + $PM_{2.5}$ hybrid	66.49	49.00	−0.280
Improvement (EffNet vs VGG)	*−15.5%*	*−16.6%*	*+0.51 abs.*

 doi:10.20944/preprints202604.2086.v1

EfficientNetB0 outperformed VGG16 in all three indicators, reducing RMSE by 15.5%, and MAE - by 16.6%. The R^2 index improved from -0.794 to -0.280, which corresponds to an absolute increase of 0.514. However, both models gave negative values of R^2, indicating that their forecasts remained less significant than with a constant average baseline. Compared to previous studies shown in the Table 1, the gap in R^2 is directly proportional to the difference in the size of the dataset: the results of studies using 3,500-15,000 images are R^2 from 0.65 to 0.94, while the current study with 1014 images gives $R^2 = -0.28$ for the best model.

4.3. True vs. Predicted AQI Scatter Plot

Figure 4 presents a scatter plot of true versus predicted AQI values of VGG16 true vs. predicted AQI in the test set. The red diagonal represents a perfect prediction. The scatter of points shows some important patterns. At low to medium AQI values (roughly 50–200), the forecasts are roughly aligned diagonally with moderate dispersion, which indicates partial signal detection within this range. For high AQI values (above 200–250), the model systematically underestimates the AQI: true values extend to 400+ on the x-axis, while the predicted values rarely exceed 200 on the y-axis. This compression effect is characteristic of regression models trained on imbalanced label distributions, where extreme values are underrepresented.

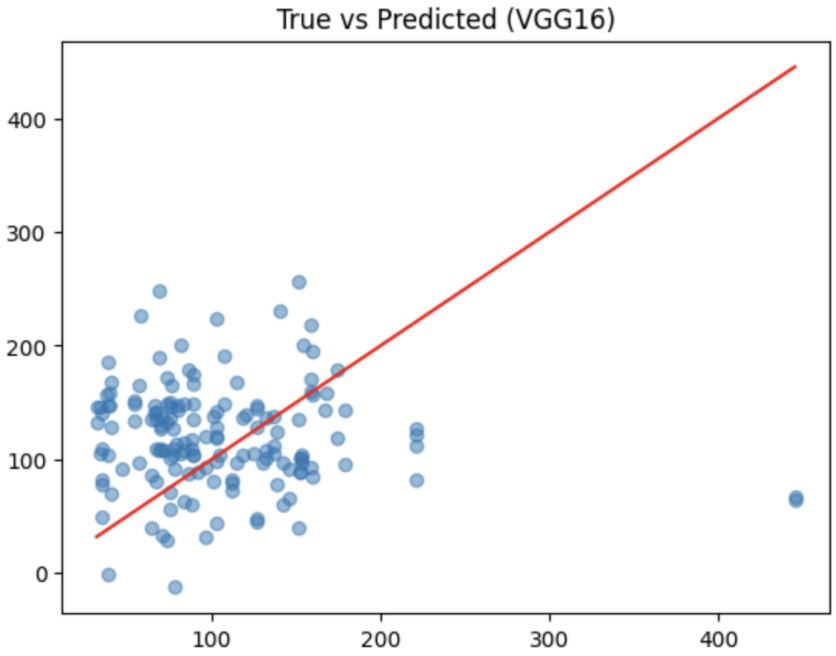

Figure 4. True vs. Predicted AQI scatter plot for the VGG16 model on the 153-sample test set. The red line represents the ideal $\hat{y} = y$ diagonal. Points cluster below the diagonal at high AQI values, indicating systematic underestimation of extreme pollution episodes.

4.4. Prediction Error Distribution

In the Figure 5 a histogram of the prediction errors (true AQI − predicted AQI)for the VGG16 model is displayed. The distribution is approximately bell-shaped and centered near zero, which indicates the absence of a strong systematic bias:the model does not consistently overestimate or underestimate the entire test set. However, the distribution exhibits a pronounced right skew with a heavy tail, which is consistent with the scatter plot finding: the model underestimates severe pollution episodes, producing large positive errors near +300–400, which correspond to extreme outlier samples likely associated with exceptional pollution events not adequately represented in the training data.

 doi:10.20944/preprints202604.2086.v1

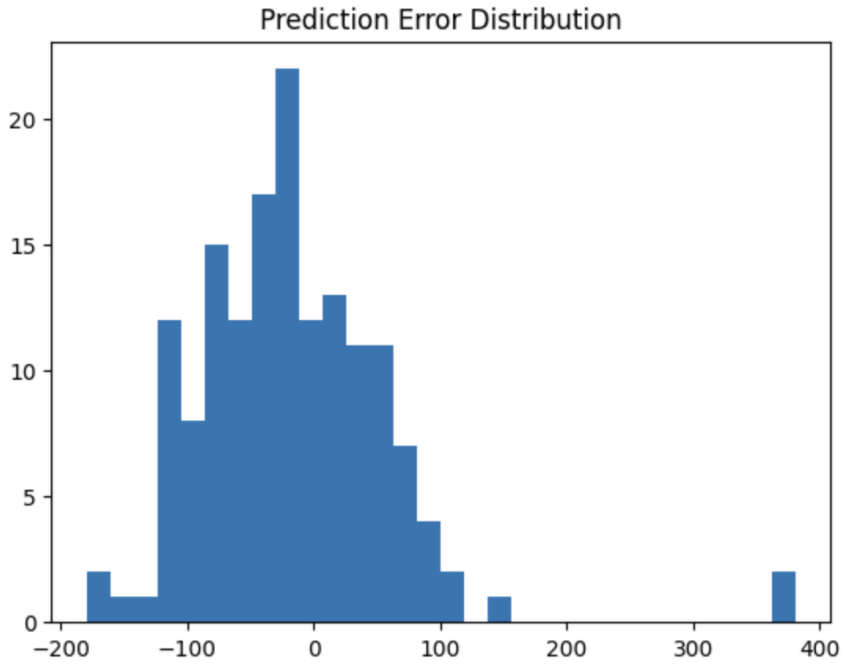

Figure 5. Prediction error distribution (true − predicted AQI) for the VGG16 model. The histogram is centred near zero but exhibits a right-skewed heavy tail, indicating systematic underestimation of high-AQI samples and the presence of extreme outlier errors.

5. Discussion

5.1. Interpreting the Negative R^2 Values

Negative R^2 values indicate that the current models are not yet suitable for reliable use. However, this does not mean the absence of a learned structure. As shown in Figure 4, the VGG16 model captures a partial prediction signal in the low to medium AQI range, with the predictions roughly aligned along the diagonal. In contrast, performance deteriorates significantly at high AQI values where the model systematically underestimates the pollutant load.

The negative R^2 is mainly caused by large errors in these samples with high contamination, which are both underrepresented in the data set and more difficult to derive from visual features alone. This imbalance between moderately predictable samples and extreme outliers leads to a general deterioration of the R^2 metric despite partial learning of the underlying patterns. It should be noted that this partial signal is a meaningful finding in itself, since the models were trained on only 710 samples under completely uncontrolled conditions.

5.2. Dataset Scale as the Primary Bottleneck

The fundamental limiting factor is the size of the dataset. For 710 training samples, the models have to calibrate the display of a multidimensional image to a scalar image in an environment with all indirect regression. Compared to the corresponding literature (Table 1), the dataset in this study is about 5 to 15 times smaller than the dataset that reaches positive values of R^2. [16] used more than 8,000 samples and received R^2= 0.71; [17] used more than 15,000 samples and received R^2= 0.78. Data-driven training will mitigate but not eliminate this limitation: even with ImageNet functions, 710 counts are not enough to reliably calibrate the display under heterogeneous real-world conditions from visual atmospheric signals to a noisy AQI scalar signal.

5.3. Contextualization of Indicators In Comparison with Previous Work

Although both models give negative values of R^2, the result is consistent with what the comparative literature predicts at this scale of data. Each study in the Table 1 reports on $R^2 \geq 0.65$ trained on 3,500–15,000 pairs [15–17,19], which is about 5-15 times more than 710 samples available here. Mondal et al. [20], who worked under directly comparable conditions in Dhaka (∼1,000 images, sparse

 doi:10.20944/preprints202604.2086.v1

monitoring), similarly observed a limited amount of prediction accuracy, providing independent verification that our results reflect the limitations of the dataset and infrastructure, and not a simulation failure.

Partial signal detection in the AQI range of 50-200 (Figure 4) confirms that networks extract genuine atmospheric characteristics rather than predicting the average for training. The consistent reduction in RMSE by 15.5% and MAE by 16.6% of EfficientNetB0 over VGG16 across all three metrics additionally eliminates random fluctuations and establishes a reproducible architecture rating for future work in this region.

5.4. Label Noise from Sensor-Image Geographic Mismatch

The AQI values are obtained from the nearest available monitoring station, which can be geographically removed from the place where the photo was taken. Urban pollution exhibits strong spatial heterogeneity on sub-kilometer scales [8], meaning that the AQI recorded at a station can systematically distort the level of pollution visible in an image that is several hundred meters or more away. This label noise cannot be reduced in view of the existing monitoring infrastructure in Bishkek and represents a fundamental upper limit for the achievable regression accuracy. No architectural improvement alone can overcome this limitation.

5.5. Visual Confounders: Illumination and Scene Diversity

Scatter plot in Figure 4 indicates that even samples with the same true AQI values can give very different predictions and vice versa. The main factor contributing to this variance is uncontrolled lighting. Visual signals correlated with AQI - cloud cover opacity, loss of contrast, changed color of the sky - react very sensitively to lighting conditions. Photos taken in bright daylight can look sharper and more contrasting than images of a cloudy morning taken with the same AQI values, which leads to an inconsistent display of object labels. In contrast to the controlled camera settings used in studies such as [18], this data set was collected under completely uncontrolled real conditions without temporal stratification or standardization of the camera angle.

5.6. Right-Skewed Error Distribution: Implications

The error distribution offset to the right, which is shown in the figure 5, is of particular importance for online provision. The model systematically underestimates episodes of high pollution - the very events that are most important for public health. From the point of view of risk management, an instrument that underestimates the level of severe contamination is less safe than an instrument that overestimates it, since incorrectly underestimated indicators may not entail medical recommendations or protective measures. Therefore, in any future implementation of an image-based AQI assessment under real conditions, this directional deviation must be precisely taken into account, for example by a class-based sample, an adjustment to the loss of focus or a special calibration with threshold correction.

5.7. EfficientNetB0 vs. VGG16: Architectural Interpretation

The continued superiority of EfficientNetB0 over all indicators (Table 4) reflects its complex scale design [5]: a coordinated scaling of depth, grid width and input resolution leads to a higher element density per parameter. In data-restricted transfer learning settings, more efficient representation compression is particularly advantageous, since the model requires fewer examples to adapt its pre-trained functions to a new domain. The inverted residual blocks of EfficientNetB0 and the consideration of compression and excitation channels additionally give the model the opportunity to selectively weigh the characteristic channels that are most suitable for atmospheric conditions, an inductive distortion that fits well with the subtask for haze detection implied in the AQI regression.

The slight improvement observed during the fine-tuning of Phase 2 for EfficientNetB0 (loss of validation from at best 0.6748 to 0.6727) suggests that the backbone has already extracted almost optimal domain-relevant functions in phase 1 and that further thawing may lead to an overfitting of a limited set of training functions. For VGG16, fine tuning led to a slightly greater improvement (MAE

 doi:10.20944/preprints202604.2086.v1

validation from 1.68 to 1.45 at best), which suggests that the older, less efficient features of VGG16 had more options for configuring the domain - corresponding to a less efficient architecture.

5.8. Recommendations for Future Work

Recent architectural achievements point to further areas of the future work. Hardini et al. [23] has shown that ensemble models that combine visual converters (ViT) with convolution networks provides an excellent AQI estimation under conditions of large amounts of data, which confirms the recommendation to study basic transformer-based systems such as ViT or change the transformer in future versions of this work. Furthermore, Aslam et al. [24] showed that a video-based evaluation with structured state space models (Mamba) reach R^2 =0.92 for AQI, by using time dependencies between successive frames, this is an approach that can be adapted to the infrastructure with fixed cameras in Bishkek, when the time series image collection is included in the data pipeline. Finally, the Grad-CAM interpretability framework demonstrated by Utomo et al. [21] provides a useful tool to determine which visual characteristics of the atmosphere – opacity, horizon contrast, color the sky temperature is the most informative for AQI regression, which will strengthen both scientific validity and practical value, the possibility of using future image-based surveillance systems in Central Asia.

Several directions can improve the results in the future. Firstly, the expansion of the data set to at least 3000 to 5000 samples, combined with controlled time stratification (morning, day, evening, season) and improved spatial diversity in the districts of Bishkek will probably have the greatest impact.

Secondly, the combination of additional sensors, including meteorological variables such as temperature, humidity and wind conditions, can partially compensate for the indirect nature of the visual assessment of AQI, which is consistent with the previous results of [18].

Thirdly, the use of more powerful backbone architectures such as EfficientNetB4 or Vision Transformers (ViT) can improve the extraction of subtle atmospheric features through a more expressive representation.

Finally, error-based learning strategies, including asymmetric loss functions or quantitative regression, can reduce the systematic undervaluation of high-efficiency events.

In addition to technical improvements, this study has revealed an important structural limitation. The effectiveness of image-based AQI assessment in Bishkek is significantly limited by the density and spatial distribution of terrestrial monitoring stations. Improved availability of cost-effective sensors, especially in low-maintenance areas, would improve the quality of the labels and enable more accurate room modeling.

6. Conclusion

This study presented the first systematic assessment of the knowledge transfer process for image-based air pollution assessment in Bishkek, Kyrgyzstan. Two hybrid models combining VGG16 and EfficientNetB0 with scalar input $PM_{2.5}$, were trained on a locally collected dataset of 1,014 image-AQI pairs. EfficientNetB0 outperformed VGG16 in all three indicators (RMSE 66.49 vs. 78.71; MAE 49.00 vs. 58.78), and both models received a partial forecast signal in the AQI range from low to medium. Negative values of R^2 reflect the limitations of the dataset and infrastructure, rather than an absence of learned structure, consistent with comparable studies at this data scale [20]. In future work, priority should be given to expanding the dataset, at least to the level of 3,000–5,000 sampling, combining meteorological metadata, and closer sensor coverage of ground-based information in Bishkek areas.

Acknowledgments: The authors thank the Ala-Too International University (AIU), Bishkek, for the academic support and appreciate the guidance of the academic supervisor. The computing resources used in this work were provided by Google Colaboratory (NVIDIA T4 GPU).

Conflicts of Interest: The authors declare no conflicts of interest.

 doi:10.20944/preprints202604.2086.v1

Data Availability Statement: The dataset of 1,014 image–AQI pairs collected in Bishkek is publicly available on Kaggle at hhttps://www.kaggle.com/datasets/mokhammadparvanivafa/urban-images-of-air-pollution-in-kyrgyzstan.

References

1. World Health Organization. WHO global air quality guidelines: Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Technical report, World Health Organization, Geneva, Switzerland, 2021.
2. Landrigan, P.J.; Fuller, R.; Acosta, N.J.R. The lancet commission on pollution and health. *Lancet* **2017**, *391*, 462–512.
3. United Nations Children's Fund. Ambient $PM_{2.5}$ air pollution in Bishkek: Key messages. Technical report, UNICEF, Bishkek, Kyrgyzstan, 2023.
4. World Bank. The global health cost of air pollution. Technical report, World Bank Group, 2023.
5. Tan, M.; Le, Q.V. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019, pp. 6105–6114.
6. GBD 2021 Risk Factors Collaborators. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations for the global burden of disease study 2021. *Lancet* **2024**, *403*, 2162–2203.
7. Maji, K.J.; Arora, M.; Dikshit, A.K. Burden of disease attributed to ambient $PM_{2.5}$ and PM_{10} exposure in 130 cities in China. *Environment International* **2021**, *123*, 105–115.
8. Kumar, P.; Omidvarborna, H.; Bhattacharya, S. The rise of low-cost sensing for managing air pollution in cities. *Nature Reviews Earth & Environment* **2021**, *2*, 196–212.
9. Liu, B.; Xu, M.; Ji, X. Low-cost outdoor air quality monitoring and source identification by dual-channel convolutional neural network on a raspberry pi. *Sensors* **2023**, *23*, 1320.
10. Li, H.; Ge, Y.; Liu, M. Estimating ground-level $PM_{2.5}$ with extra-trees across China. *International Journal of Environmental Research and Public Health* **2022**, *19*, 4084.
11. Zhao, J.; Zhao, X.; Xu, H. Air quality prediction using multimodal data and transfer learning. *IEEE Access* **2022**, *10*, 12345–12358.
12. Gulia, S.; Nagendra, S.M.S.; Khare, M.; Khanna, I. Urban air quality management: A review. *Atmospheric Pollution Research* **2022**, *13*, 101286.
13. Burmachach, N.; Isaev, R.; Gimaletdinova, G. Predicting passenger car prices with machine learning models. Preprints, 2024. https://doi.org/10.20944/preprints202412.0849.v1.
14. Esenalieva, G.; Khan, M.T.; Ermakov, A.; Tursunbekova, E.T. Real-time sign language recognition. *Alatoo Academic Studies* **2024**, *2024*, 165–174.
15. Kow, P.Y.; Hsia, I.W.; Chang, L.C.; Chang, F.J. Real-time image-based air quality estimation by deep learning neural networks. *Journal of Environmental Management* **2022**, *307*, 114555.
16. Zhang, Q.; Fu, F.; Tian, R. A deep learning and image-based model for air quality estimation. *Science of the Total Environment* **2020**, *724*, 138178.
17. Xue, Y.; Wang, Y.; Zhang, D. Image-based air quality analysis using deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* **2023**, *61*, 1–14.
18. Dong, L.; Li, L.; Jiang, M. Estimating $PM_{2.5}$ concentration from a single image using convolutional neural networks. *Remote Sensing* **2021**, *13*, 2148.
19. Wang, X.; Wang, M.; Liu, X.; Mao, Y.; Chen, Y.; Dai, S. Surveillance-image-based outdoor air quality monitoring. *Environmental Science and Ecotechnology* **2024**, *18*, 100319.
20. Mondal, J.J.; Islam, M.; Islam, R.; Rhidi, N.; Newaz, S.; Manab, M.A.; Islam, A.B.M.A.A.; Noor, J. Uncovering local aggregated air quality index with smartphone captured images leveraging efficient deep convolutional neural network. *Scientific Reports* **2024**, *14*, 1320.
21. Utomo, S.; Rouniyar, A.; Jiang, G.H.; Chang, C.H.; Tang, K.C.; Hsu, H.C.; Hsiung, P.A. Air quality prediction from images in Indonesia: enhancing model explainability through visual explanation with AQI-Net and Grad-CAM. *Environmental Data Science* **2024**, *3*, e25.
22. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the Proceedings of the International Conference on Learning Representations, 2015.

 doi:10.20944/preprints202604.2086.v1

23. Hardini, A.; Kusuma, P.; Dewi, R. An ensemble deep learning approach for air quality estimation in Delhi, India. *Earth Science Informatics* **2024**, *17*, 891–906.
24. Aslam, N.; Khan, I.U.; Mirza, S. Low-cost video-based air quality estimation system using structured deep learning with selective state space modeling. *Environment International* **2025**, *198*, 109012.