Physics-Constrained Lightweight Neural Networks for Calibrated Smog-Level Classification

Sharon Christa

Mansi Bhonsle

School of Computing
MIT Art Design and Technology University
Pune, India

Abstract

Air pollution is a major environmental and health challenge, with fine particulate matter ($PM_{2.5}$) responsible for millions of premature deaths each year. Ground-based sensors provide reliable air quality measurements but are costly and sparsely distributed, limiting large-scale coverage. As a complementary approach, image-based methods using street-level photographs have emerged, though most rely on large, computationally heavy networks that neglect physical principles of haze formation and often produce poorly calibrated outputs. This paper introduces a lightweight, physics-informed, and calibrated framework for smog-level classification from street-level imagery. The backbone model is MobileNetV3-Small, designed for efficient CPU inference. To better recognize minority pollution categories, physics-informed regularization is applied by embedding atmospheric scattering cues—contrast, sharpness, and dark channel prior—into the loss function. Reliability is enhanced through test-time augmentation and temperature scaling, while interpretability is addressed using Grad-CAM visualizations and monotonic physics-feature trends. Experiments on the Smartphone-Based Air Pollution Image Dataset (SAPID) show that the proposed Physics v2 model achieves 86.5% test accuracy and a macro-F1 score of 0.835, surpassing the baseline MobileNetV3-Small (81.1%, 0.756 macro-F1). The framework also operates in real time on CPU hardware at over 100 FPS, with an Expected Calibration Error (ECE) of 0.071. These results demonstrate the potential of combining lightweight architectures, physics priors, and calibration techniques to deliver accurate, interpretable, and deployable vision systems for low-cost urban air quality monitoring.

1 Introduction

Air pollution is among the most significant global health risks, with approximately seven million premature deaths annually linked to fine particulate matter and other pollutants [14]. Chronic exposure to $PM_{2.5}$ and PM_{10} is associated with cardiovascular and respiratory diseases, as well as impaired child lung development [2]. Rapid urbanization and industrialization, especially in low- and middle-income countries, exacerbate air quality challenges where dense monitoring infrastructure is lacking.

Ground-based stations provide precise pollutant measurements but remain expensive to install and maintain, leading to sparse coverage and limiting fine-grained assessments [8]. To overcome this, recent research has explored low-cost alternatives leveraging ubiquitous sensing modalities such as smartphones and public cameras. Street-level imagery captures visual indicators of pollution, including haze and reduced visibility, making it a promising complementary data source.

Early computer vision methods employed handcrafted features such as edge sharpness and color attenuation. More recent approaches apply deep learning, directly regressing particulate levels from images [8, 11]. Despite promising results, three key limitations remain: (i) reliance on large models like ResNet and VGG that require GPUs, hindering deployment in resource-constrained settings; (ii) neglect of atmospheric scattering principles underlying haze formation; (iii) class imbalance in critical but underrepresented categories such as “Unhealthy for Sensitive Groups” and “Very Unhealthy”; and (iv) overconfident predictions, underscoring the need for calibration [4].

Lightweight networks such as MobileNetV3 [6] offer efficiency for edge devices, while physics-informed learning introduces domain priors into training. Atmospheric cues like contrast, sharpness, and dark channel priors are particularly relevant to haze modeling [5, 13]. Calibration methods (e.g., temperature scaling) enhance reliability [4], and interpretability techniques like Grad-CAM [18] provide transparency.

This study proposes a lightweight, physics-informed, and calibrated framework for smog-level classification using MobileNetV3-Small. Evaluated on the Smartphone-based Air Pollution Image Dataset (SAPID), the approach integrates physics priors and calibration to improve recognition of minority classes, ensure reliable predictions, and enable efficient CPU-only deployment.

2 Related Work

Compact CNNs have enabled efficient on-device perception. MobileNetV2 introduced inverted residuals with linear bottlenecks and depthwise separable convolutions [17], while MobileNetV3 added h-swish activation, squeeze-and-excitation, and hardware-aware NAS for state-of-the-art accuracy under latency constraints [6]. These remain strong baselines for CPU-only deployment. Adverse-weather vision builds on the atmospheric scattering model, where observed radiance combines attenuated scene radiance and airlight [13]. The dark channel prior (DCP) became a seminal dehazing heuristic [5], and benchmarks such as RESIDE [10] established standardized evaluation. Air-quality estimation evolved from handcrafted cues to CNN-based regression/classification of $PM_{2.5}$/AQI from street or surveillance imagery, sometimes extended with temporal modeling (CNN–LSTM) [22]. Practical deployments for real-time regression [9] and hybrid pipelines with satellites further highlight its applicability.

PINNs embed governing equations as constraints to improve physical fidelity [16]. For haze/smog vision, cues such as contrast, Laplacian sharpness, and DCP align with scattering physics and are increasingly used as priors or regularizers [7]. Deep models are often miscalibrated. Guo *et al.* [4] showed that temperature scaling significantly improves reliability without reducing accuracy. In environmental monitoring, calibrated confidence estimates are critical for thresholding and decision support. RESIDE [10] supports dehazing evaluation, while HVAQ links images with pollutant and meteorological data across cities [1]. SAPID [24] provides smartphone photos grouped into five EPA AQI categories, enabling lightweight classification studies. Deployment faces dataset shift across cameras, cities, seasons, and co-occurring weather [15]. Classical domain adaptation aligns source and target via CORAL [19] or adversarial learning [3]. Surveys [21, 23] document strong gains, while recent test-time adaptation dispenses with source data and adapts using entropy minimization (TENT) [20] or source-free SHOT [12]. Such lightweight adaptation is promising for environmental vision where labels are scarce.

Across lightweight CNNs, image-based haze/AQ estimation, physics-informed learning, calibration, and domain adaptation, several gaps remain salient for deployment in resource-constrained urban monitoring since many models assume GPU availability and large backbones; fewer studies report accuracy–latency trade-offs for CPU-only inference suitable for smart phones or embedded devices [6]. Learning pipelines often remain purely data-driven; the literature shows fewer examples where atmospheric scattering cues (contrast, sharpness, dark-channel statistics) are explicitly enforced during training in a way that improves minority-class recognition. Image-based AQ estimation systems rarely report calibration metrics; yet calibrated confidence is critical for thresholding and alerting in environmental applications [4]. Datasets such as SAPID exhibit severe class imbalance, with under-represented high-severity categories. Robust learning under such imbalance, coupled with interpretability, remains under-explored at the edge. Cross-region generalization and test-time adaptation are underutilized in AQ-from-images, despite clear distribution shifts across cameras, cities, and seasons [12, 20, 23].

These observations indicate a research gap for a *lightweight, physics-informed, and calibrated* air-quality classifier that performs on CPU-class hardware, improves minority-class recognition, and remains robust under realistic distribution shifts via simple test-time procedures. Such a design directly addresses operational constraints in low-cost, scalable urban monitoring.

3 Research Methods

The Smartphone-Based Air Pollution Image Dataset (SAPID) [24] is used as the primary benchmark. It consists of 492 street-level images annotated with five air quality categories following the US Environmental Protection Agency (EPA) Air Quality Index (AQI): Good, Moderate, Unhealthy for Sensitive Groups (USG), Unhealthy, and Very Unhealthy. As shown earlier (Table **??**), the dataset is highly imbalanced, with only 32 samples in the USG class and 40 in Very Unhealthy, in contrast to 188 in the Moderate class. Images are resized to 224×224 and normalized using ImageNet mean and standard deviation. Data augmentation includes random flips, rotations, and color jittering to improve generalization [24]. The baseline classifier is based on MobileNetV3-Small [6], which uses depthwise separable convolutions, inverted residuals, and squeeze-and-excitation blocks with h-swish activations. For an input image $x \in R^{3\times H\times W}$, the network extracts features $f(x) \in R^d$, which are passed to a fully connected classifier:

$$z = Wf(x) + b, \quad \hat{y} = \text{softmax}(z),$$

where $z \in R^C$ are the logits, $C = 5$ is the number of AQI classes, and $\hat{y}$ is the predicted probability distribution. The baseline is trained using class-weighted cross-entropy loss:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{C} w_i y_i \log \hat{y}_i,$$

where y is the one-hot ground truth vector and w_i are class weights to address dataset imbalance.

3.1 Physics-Informed Model

To integrate atmospheric scattering priors [5, 13], three physics-inspired features are extracted from each input image:

1. **Contrast ($C(x)$):** Standard deviation of pixel intensities or global contrast tensor, expected to decrease with pollution severity.

2. **Sharpness ($S(x)$):** Laplacian variance measuring edge clarity, also decreasing under haze.

3. **Dark Channel Prior ($D(x)$):** Defined as

$$D(x) = \min_{c \in \{R,G,B\}} \left(\min_{u \in \Omega(x)} I_c(u) \right),$$

   where $\Omega(x)$ is a local patch around pixel x, and $I_c(u)$ is the intensity in color channel c. The DCP increases under heavier smog.

Let $\phi(x) = [C(x), S(x), D(x)]$ be the physics feature vector. The model enforces monotonic consistency by penalizing violations of the expected order across pollution categories. For two samples (x_i, y_i) and (x_j, y_j) with $y_i < y_j$ (less polluted vs. more polluted), the physics-informed ranking loss is defined as:

$$\mathcal{L}_{physics} = \sum_{i,j} \Big(\max\big(0, (C(x_j) - C(x_i))\big) + \max\big(0, (S(x_j) - S(x_i))\big) + \max\big(0, (D(x_i) - D(x_j))\big) \Big).$$

This enforces decreasing contrast and sharpness, and increasing dark channel prior, with rising pollution severity.

The total objective is:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \lambda_{phys} \mathcal{L}_{physics},$$

where $\lambda_{phys} = 0.2$ balances classification and physics constraints.

Neural networks are often miscalibrated, producing overconfident predictions [4]. Logits z are calibrated using temperature scaling:

$$\hat{y}_i^{cal} = \frac{\exp(z_i / T)}{\sum_{j=1}^{C} \exp(z_j / T)},$$

where $T > 0$ is a learned temperature parameter optimized on the validation set. A perfectly calibrated model satisfies

$$P(Y = y \mid \hat{p}) = \hat{p}, \quad \forall \hat{p} \in [0, 1].$$

To improve robustness, test-time augmentation (TTA) generates K transformations $\{t_k(x)\}_{k=1}^{K}$ for each input, and predictions are averaged:

$$\hat{y}_{TTA} = \frac{1}{K} \sum_{k=1}^{K} \hat{y}(t_k(x)).$$

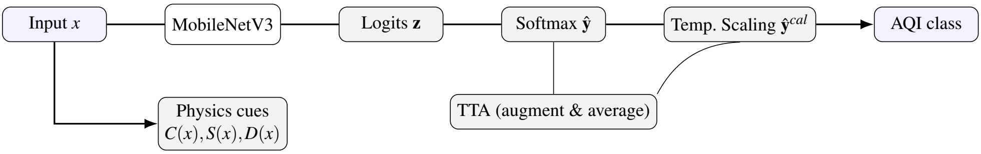

Input x
MobileNetV3
Logits $\mathbf{z}$
Softmax $\hat{\mathbf{y}}$
Temp. Scaling $\hat{\mathbf{y}}^{cal}$
AQI class
Physics cues $C(x), S(x), D(x)$
TTA (augment & average)

Figure 1: Architecture overview

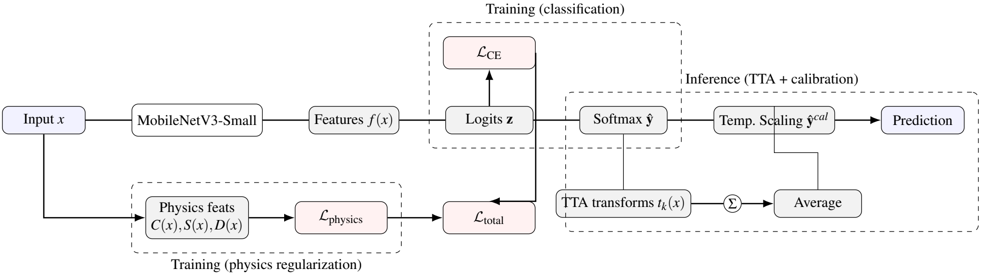

Training (classification)
$\mathcal{L}_{\text{CE}}$
Inference (TTA + calibration)
Input x
MobileNetV3-Small
Features $f(x)$
Logits $\mathbf{z}$
Softmax $\hat{\mathbf{y}}$
Temp. Scaling $\hat{\mathbf{y}}^{cal}$
Prediction
Physics feats $C(x), S(x), D(x)$
$\mathcal{L}_{\text{physics}}$
$\mathcal{L}_{\text{total}}$
TTA transforms $t_k(x)$
Σ
Average
Training (physics regularization)

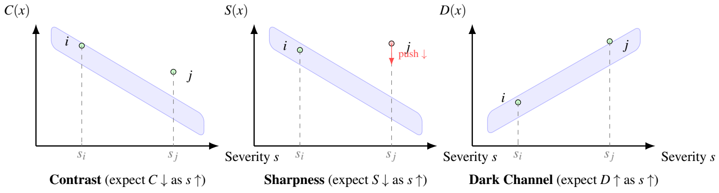
Figure 2: End-to-end pipeline

Physics loss mechanics for pairwise ranking is depicted in Figure 3. For two samples with severities $s_i < s_j$, contrast C and sharpness S are expected to decrease, while dark channel D is expected to increase with severity. A hinge margin δ penalizes violations, forming the physics-informed component of the total objective.

The models are trained in PyTorch using the AdamW optimizer with weight decay 10^{-4} and an initial learning rate of 3×10^{-4}, following a cosine decay schedule. A batch size of 32 is used. Early stopping is triggered if validation macro-F1 does not improve for six consecutive epochs. To address dataset imbalance, class-weighted cross-entropy with label smoothing ($\varepsilon = 0.05$) is employed. Gradient clipping (ℓ_2 norm capped at 1.0) prevents instability. Performance is evaluated using accuracy and macro-F1, the latter being particularly important under imbalanced data distributions. In addition, per-class precision, recall, and average precision (AP) are reported. Reliability of predicted probabilities is measured using Expected Calibration Error (ECE) [4]. Model efficiency is evaluated in terms of parameter count, file size, CPU latency (milliseconds per image), and frames per second (FPS). Interpretability is assessed qualitatively with Grad-CAM [18] and quantitatively with physics-feature alignment trends.

4 Experimental Setup

All experiments were conducted in Google Colab with access to an NVIDIA Tesla T4 GPU (16 GB) when available and CPU-only mode otherwise. The model architecture and training routines were implemented in PyTorch 2.1, using the Torchvision model zoo for MobileNetV3-Small initialization [6]. Training and evaluation pipelines were executed under Python 3.12, with supporting libraries including NumPy, Pandas, and scikit-learn. Grad-CAM visualizations were generated using the TorchCAM library. The SAPID dataset [24] was split into training (70%), validation (15%), and test (15%) sets, ensuring class-stratified sampling to preserve distribution across splits. The validation set was used for hyperpa-


$C(x)$
i
j
s_i
s_j
Severity s
Contrast (expect $C\downarrow$ as $s\uparrow$)
$S(x)$
i
j
push ↓
s_i
s_j
Severity s
Sharpness (expect $S\downarrow$ as $s\uparrow$)
$D(x)$
i
j
s_i
s_j
Severity s
Dark Channel (expect $D\uparrow$ as $s\uparrow$)

Pairwise ranking loss for $s_i < s_j$:
$\mathcal{L}_{\text{phys}} = \max\big(0, C(x_j) - C(x_i) + \delta\big) + \max\big(0, S(x_j) - S(x_i) + \delta\big) + \max\big(0, D(x_i) - D(x_j) + \delta\big).$
Margin $\delta > 0$ enforces monotonic ordering; total objective $\mathcal{L}_{\text{total}} = \mathcal{L}_{CE} + \lambda_{\text{phys}}\mathcal{L}_{\text{phys}}$.

Figure 3: Physics-informed pairwise ranking

rameter tuning, early stopping, and calibration (temperature scaling). The final test set was reserved strictly for performance reporting. All images were resized to 224 × 224 pixels and normalized with ImageNet mean and standard deviation values. To increase robustness and mitigate overfitting, training augmentations included random horizontal flips and random rotations (±15°), random brightness, contrast, and saturation jitter, random cropping and scaling. For inference, standard resizing was applied, and test-time augmentation (TTA) was used to generate multiple crops and scales per image, with averaged predictions. Models were trained using the AdamW optimizer with an initial learning rate of 3×10^{-4} and cosine decay scheduling. A batch size of 32 and early stopping with patience of six epochs were employed. Gradient clipping with a maximum ℓ_2 norm of 1.0 stabilized optimization. To address dataset imbalance, class-weighted cross-entropy loss was combined with label smoothing ($\varepsilon = 0.05$). For the physics-informed variant, a regularization weight of $\lambda_{phys} = 0.2$ was used for the physics-based loss component. Performance was assessed using overall accuracy and macro-F1 score, the latter being critical for imbalanced datasets. Class-wise precision, recall, and average precision (AP) were also reported. Calibration quality was evaluated via Expected Calibration Error (ECE) following [4]. Model interpretability was analyzed through Grad-CAM heatmaps [18] and physics-feature trend plots (contrast, sharpness, dark channel prior). Efficiency metrics included parameter count, model file size, CPU latency, and frames per second (FPS). The following configurations were compared: 1) MobileNetV3-Small with standard cross-entropy training. 2) Initial physics-informed model with basic regularization. 3) Improved physics-informed model with tuned λ_{phys} and ranking constraints. 3) Incorporating TTA and temperature scaling for reliability. 4) Weighted averaging of baseline and physics-informed predictions. This setup ensures a fair ablation study and highlights the contributions of physics-informed regularization and calibration.

5 Results and Discussion

Figure 4 shows the SAPID dataset distribution, highlighting significant class imbalance. Minority categories (USG and Very Unhealthy) contain fewer than 50 samples each, motivating the use of class weighting and physics-informed regularization.

The baseline MobileNetV3-Small achieved 81.1% test accuracy and 0.756 macro-F1 with TTA. Figure 5 presents the confusion matrix and one-vs-rest precision-recall (PR) curves. While performance was strong for "Good" and "Moderate", the model misclassi-

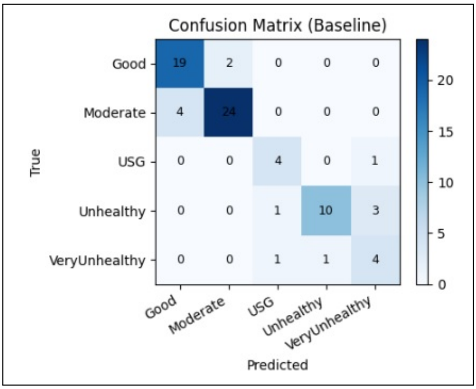
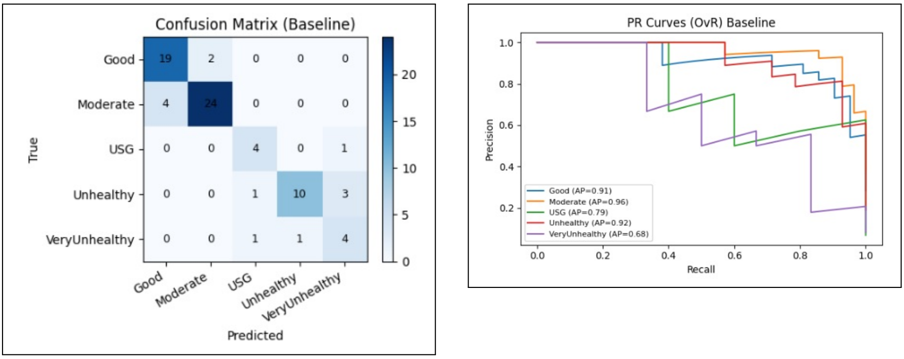


Figure 4: Class distribution of SAPID dataset across five AQI categories.



(a) Confusion Matrix (Baseline)

(b) PR Curves (Baseline)

Figure 5: Baseline results: (a) Confusion Matrix; (b) PR curves

fied minority categories (USG and Very Unhealthy).

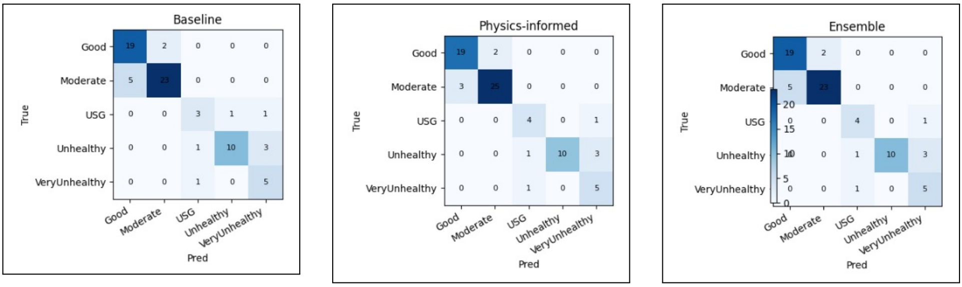
The Physics v2 model improved recognition of minority categories by enforcing monotonic trends in haze-sensitive features. It achieved 85.1% accuracy and 0.804 macro-F1 with TTA. With calibration, performance further improved to 86.5% accuracy and 0.835 macro-F1. Figure 6 compares confusion matrices of baseline, physics-informed, and ensemble models.

To validate interpretability, physics features (contrast, sharpness, dark channel prior) were analyzed against predicted classes. As shown in Figure 7, contrast and sharpness decrease with increasing pollution severity, while the dark channel prior increases. These trends are consistent with atmospheric scattering theory.

Figure 8 presents Grad-CAM visualizations for sample test images. The model consistently attends to haze-heavy regions of the sky and building outlines, confirming that decisions align with haze-relevant image regions rather than spurious background features.

Temperature scaling improved probability calibration. Figure 9 shows the reliability diagram of Physics v2 after calibration, with an Expected Calibration Error (ECE) of 0.071. This demonstrates alignment between predicted confidence and observed accuracy.

Table 1 compares efficiency metrics. Both baseline and physics-informed models remain lightweight, with ∼1.5M parameters, ∼6 MB file size, and CPU inference speeds exceeding 100 FPS. Physics-informed modifications incur negligible additional cost.



(a) Baseline (b) Physics v2 (c) Ensemble

Figure 6: Confusion matrices comparing Baseline, Physics v2, and Ensemble models.

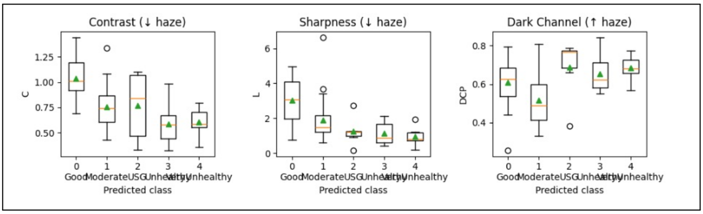
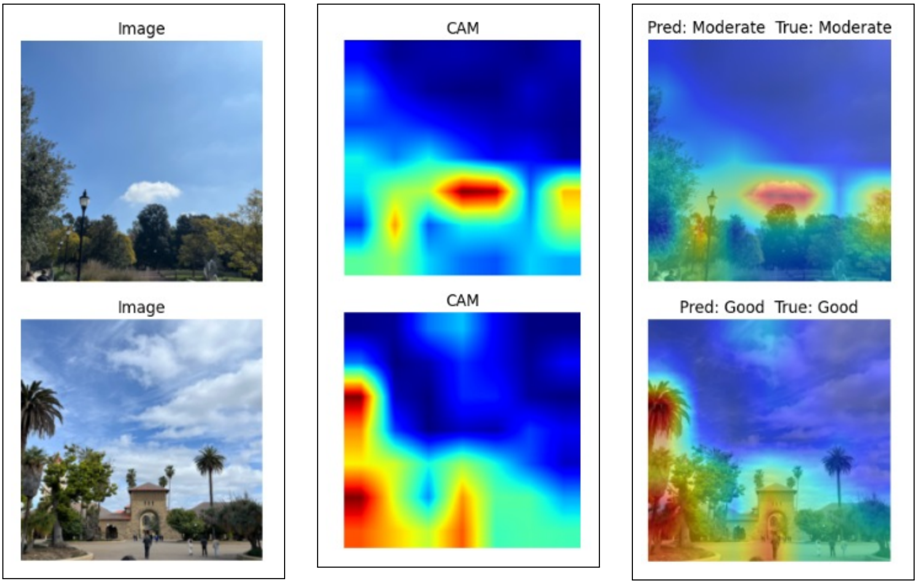


Figure 7: Boxplots of physics features across predicted classes.



(a) Input Image (b) Grad-CAM Heatmap (c) Overlay

Figure 8: Grad-CAM visualizations.

Model	Params (M)	File Size (MB)	CPU Latency (ms)	FPS
Baseline (v3-small)	1.52	6.23	9.16	109.2
Physics v2 (v3-small)	1.52	6.23	8.99	111.2

Table 1: Efficiency comparison between baseline and Physics v2 models.

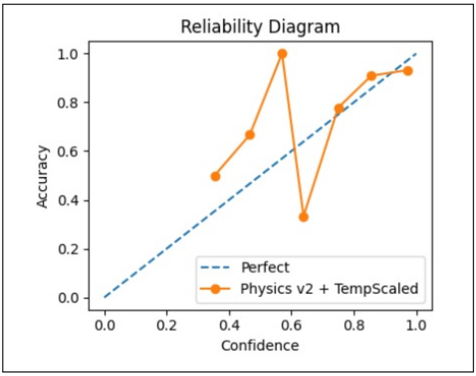


Figure 9: Reliability diagram for Physics v2 + temperature scaling.

Model	Test Accuracy	Macro-F1
Baseline (TTA)	0.811	0.756
Physics v2 (TTA)	0.851	0.804
Physics v2 + TTA + Temp Scaling	**0.865**	**0.835**
Ensemble (equal weights, TTA)	0.824	0.788

Table 2: Final results summary on SAPID test set

Table 2 reports the final comparison of baseline, physics-informed, and calibrated models. The Physics v2 model with TTA and temperature scaling achieved the best performance with 86.5% accuracy and 0.835 macro-F1.

6 Conclusion and Future Work

This study introduced a lightweight, physics-informed, and calibrated framework for smog-level classification from street-level imagery. Built on MobileNetV3-Small, the model integrates atmospheric scattering priors—contrast, sharpness, and dark channel statistics—into the loss function to improve recognition of minority classes in the imbalanced SAPID dataset. With test-time augmentation and temperature scaling, the Physics v2 model achieved 86.5% accuracy and a macro-F1 of 0.835, surpassing the baseline while maintaining real-time CPU performance at over 100 FPS. Grad-CAM analyses and monotonic physics-feature trends enhanced interpretability, and calibration reduced Expected Calibration Error to 0.071, ensuring reliable confidence estimates.

The results highlight three contributions: (1) CPU-ready lightweight models for urban air quality monitoring, (2) physics-informed constraints that improve robustness under class imbalance, and (3) calibration as a requirement for trustworthy deployment.

Future Research Directions: While promising, several extensions remain like addressing dataset shifts across regions, seasons, and devices via source-free or test-time methods [12, 20]. Combining street-level imagery with meteorological or satellite data to improve robustness. Applying semi- or weakly-supervised learning to mitigate scarcity in high-pollution categories. Benchmarking on smartphones, Raspberry Pi, and Jetson Nano for practical scalability.Exploring attribution methods and physics-driven interpretability beyond Grad-CAM. By uniting efficiency, interpretability, and calibration, this framework advances vision-based environmental intelligence and lays groundwork for scalable, low-cost air quality monitoring. Further, while evaluated on SAPID, future work will address cross-city and seasonal generalization, and integration with real-world monitoring networks.

References

[1] Zuohui Chen, Tony Zhang, Zhuangzhi Chen, Yun Xiang, Qi Xuan, and Robert P. Dick. Hvaq: A high-resolution vision-based air quality dataset. *arXiv preprint arXiv:2102.09332*, 2021. doi: 10.48550/arXiv.2102.09332. URL https://arxiv.org/abs/2102.09332.

[2] Aaron J Cohen, Michael Brauer, Richard Burnett, Heather R Anderson, Joseph Frostad, Kara Estep, Kalpana Balakrishnan, Bert Brunekreef, Lalit Dandona, Rakhi Dandona, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *The Lancet*, 389(10082):1907–1918, 2017.

[3] Yaroslav Ganin and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016. URL http://jmlr.org/papers/v17/15-239.html.

[4] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1321–1330. PMLR, 2017.

[5] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1956–1963, 2009. URL https://projectsweb.cs.washington.edu/research/insects/CVPR2009/award/hazeremv_drkchnl.pdf.

[6] Andrew Howard, Mark Sandler, Bo Chen, Weijun Wang, Yukun Chen, Mingxing Tan, Quoc V Le Chu, and Hartwig Adam. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1314–1324, 2019.

[7] Chih-Hao Hsieh, Yi-Jui Chen, Hsueh-Yi Kuo, Sheng-Yi Chen, and Yi-Hsuan Tsai. Using haze level estimation in data cleaning for single-image dehazing. *Electronics*, 12(16):3485, 2023. doi: 10.3390/electronics12163485. URL https://www.mdpi.com/2079-9292/12/16/3485.

[8] Jun Huang, Xin Yang, Zhixiong Zheng, and Min Li. Deep haze estimation with convolutional neural network for air quality monitoring from images. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1066–1071. IEEE, 2019.

[9] Pei-Ying Kow, I.-W. Hsia, Li-Chiu Chang, and Fang-Ju Chang. Real-time image-based air quality estimation by deep learning neural networks. *Journal of Environmental Management*, 307:114560, 2022. doi: 10.1016/j.jenvman.2022.114560. URL https://doi.org/10.1016/j.jenvman.2022.114560.

[10] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *arXiv preprint arXiv:1712.04143*, 2017. URL https://arxiv.org/abs/1712.04143.

[11] Chen Li, Wei Zhang, and Jian Sun. Smog level prediction from webcam images using deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 22(8):5091–5100, 2021.

[12] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 6028–6039. PMLR, 2020. URL https://proceedings.mlr.press/v119/liang20a.html.

[13] Srinivasa G. Narasimhan and Shree K. Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, 2002. doi: 10.1023/A:1016328200723. URL https://cave.cs.columbia.edu/old/publications/pdfs/Narasimhan_IJCV02.pdf.

[14] World Health Organization. Air pollution. *WHO Fact Sheet*, 2020. URL https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.

[15] Joaquin Quionero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence. *Dataset Shift in Machine Learning*. MIT Press, 2009. URL https://mitpress.mit.edu/9780262170055/dataset-shift-in-machine-learning/.

[16] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045. URL https://www.sciencedirect.com/science/article/pii/S0021999118307125.

[17] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4510–4520, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/papers/Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.pdf.

[18] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626. IEEE, 2017.

[19] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European Conference on Computer Vision (ECCV) Workshops*, pages 443–450, 2016. doi: 10.1007/978-3-319-49409-8_35. URL https://arxiv.org/abs/1607.01719.

[20] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://openreview.net/forum?id=uXl3bZLkr3c.

[21] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018. doi: 10.1016/j.neucom.2018.05.083. URL https://doi.org/10.1016/j.neucom.2018.05.083.

[22] Xiaochu Wang, Meizhen Wang, Xuejun Liu, Ying Mao, and Yang Chen. Surveillance-image-based outdoor air quality monitoring. *Environmental Science and Ecotechnology*, 18:100319, 2023. doi: 10.1016/j.ese.2023.100319. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10569950/.

[23] Garrett Wilson and Diane J. Cook. A survey of unsupervised domain adaptation for visual recognition. *ACM Transactions on Intelligent Systems and Technology*, 11(5):1–46, 2020. doi: 10.1145/3400066. URL https://dl.acm.org/doi/10.1145/3400066.

[24] Ruikuan Zhu. Smartphone-based air pollution image dataset (sapid). Mendeley Data, Version 1, 2024. URL https://data.mendeley.com/datasets/j654cspb6r/1.