



Probabilistic Machine Learning Early Warning for Urban PM_{2.5} in SEA Cities

Baharuddin Mide¹, Dhimas Tribuana^{2*}, Usman Sattar³, Dayanti⁴

¹ Computer Science Faculty, Universitas Ichsan Sidenreng Rappang, Sidrap, Indonesia.

² Informatics Management Faculty, Akademi Sekretari Manajemen Indonesia Publik, Makassar, Indonesia.

³ Politeknik Bombana, Bombana, Indonesia

⁴ Information Technology Department, Faculty of Engineering, Universitas Patria Artha, Makassar, Indonesia

*d.tribuana@gmail.com

Abstract

Air pollution, particularly fine particulate matter (PM_{2.5}), poses a critical threat to public health in rapidly urbanizing regions. Reliable early-warning systems are essential for mitigating exposure risks, yet challenges remain in cities with heterogeneous sensor coverage and event frequency. This study aimed to develop and evaluate a probabilistic, portable across cities early-warning framework for PM_{2.5} exceedances in Southeast Asia, focusing on Jakarta, Singapore, and Bangkok. Using a staged experimental design (Exp-A through Exp-E), we integrated regression-based back-casts with classification-based exceedance alerts, applied variant selection across thresholds and horizons, and validated for operational readiness through model freezing, documentation, and online simulation. Results showed that Jakarta achieved near-perfect exceedance prediction up to 6-hour horizons ($F1 \approx 0.99$), Singapore maintained strong performance at short horizons ($F1 \approx 0.91$ at 2–3 hours), while Bangkok yielded moderate but actionable signals at very short horizons ($F1 \approx 0.62$ at 1 hour). Regression components provided stable situational awareness, and online smoothing reduced false alarms by approximately 15–20% without degrading performance. The framework demonstrated that calibrated exceedance probabilities can serve as an effective basis for city-level air quality alerts, with reliability strongly influenced by data density and event prevalence. This work contributes a reproducible, transparent, and computationally efficient approach that bridges machine learning innovation with practical environmental management. The findings emphasize the importance of horizon-specific calibration and adaptive strategies, offering both theoretical insights and practical value for policymakers in urban air quality governance.

Keywords: air quality monitoring; early warning; exceedance prediction; machine learning; PM_{2.5}

How to Cite: B. Mide, D. Tribuana, U. . Sattar, and D. Dayanti, "Probabilistic Machine Learning Early Warning for Urban PM_{2.5} in SEA Cities", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 4, pp. 1144–1153, Aug. 2026, doi: [10.29207/resti.v10i7199](https://doi.org/10.29207/resti.v10i7199).
Permalink/DOI: <https://doi.org/10.29207/resti.v10i7199>

Received: September 23, 2025

Accepted: April 12, 2026

Available Online: August 23, 2026

This is an open-access article under the [CC BY 4.0 License](https://creativecommons.org/licenses/by/4.0/)

Published by [Ikatan Ahli Informatika Indonesia](https://www.iaii.or.id/)

1. Introduction

Air pollution remains a critical environmental and public health concern, with fine particulate matter (PM_{2.5}) linked to millions of premature deaths annually worldwide [1]. Urban centers in Asia face particular challenges due to rapid population growth, traffic emissions, industrial activities, and seasonal transboundary haze [2], [3]. In Southeast Asia, cities such as Bangkok, Jakarta, and Singapore frequently experience episodic PM_{2.5} exceedances that pose risks to vulnerable populations [4], [5].

Traditional air quality forecasting systems rely heavily on deterministic chemical transport models (CTMs), which are computationally intensive and sensitive to uncertain emissions inventories [6–8]. In recent years,

machine learning (ML) and deep learning approaches have emerged as effective alternatives or complements, offering data-driven adaptability and lower computational costs. Hybrid deep learning models combining temporal encoders (e.g. transformer + LSTM) have shown improved forecasting accuracy across cities [9]. Graph-based neural networks, such as GNN-LSTM hybrids, have been used for multi-station forecasting with spatial correlations in China and the Beijing-Tianjin-Hebei region [10].

However, the majority of prior work emphasizes point forecasts of concentration levels. While accurate regression is valuable, operational stakeholders often require early warning of exceedance decisions on whether air quality will surpass critical thresholds (e.g., 25 or 35 $\mu\text{g}/\text{m}^3$) within a short horizon. Exceedance

framing has gained recent traction in Europe, as shown by deep learning models forecasting intra-day exceedances [2], China through studies predicting PM_{2.5} exceedance days using machine learning [11], and in multi-regional operational systems such as urbanAQF [12]. These works suggest that probabilistic approaches (e.g., forecasting probabilities of threshold exceedance) offer better calibration for alerting and decision support, especially under data uncertainty.

Despite these advances, there remains a literature gap in Southeast Asia. Few studies have systematically evaluated probabilistic ML methods for PM_{2.5} exceedance prediction across heterogeneous urban networks. There have been no probabilistic studies across Southeast Asian cities with heterogeneous sensor constraints. Monitoring infrastructures vary significantly: Singapore operates a relatively dense and stable monitoring network, whereas Jakarta and Bangkok continue to face challenges with limited coverage and intermittent sensor data [13-15]. This heterogeneity limits the portability of complex spatio-temporal models and highlights the need for pragmatic, transferable pipelines that can function under data sparsity and irregularities.

To address this gap, our study investigates probabilistic machine learning for early warning of urban PM_{2.5} exceedances in Bangkok, Jakarta, and Singapore. Building on recent literature that emphasizes exceedance framing [2], hybrid temporal deep learning models [9], graph-based neural architectures [10], and hybrid forecasting approaches [7], we frame the task as a probabilistic classification problem: given past air quality and meteorological features, estimate the probability that PM_{2.5} will exceed a defined threshold within the next six hours. Beyond offline evaluation, we implement an online monitoring simulation (Exp-E) to test operational deployment, logging alerts and stability metrics at the city level.

The contributions of this paper are threefold. First, this study provides one of the first systematic evaluations of probabilistic machine learning-based PM_{2.5} exceedance early warning across multiple Southeast Asian cities using a unified experimental framework. While prior studies largely focus on concentration forecasting or single-city analyses, comparative probabilistic exceedance modeling across heterogeneous Southeast Asian monitoring environments remains limited. Second, we introduce an integrated framework that combines regression-based situational awareness with calibrated probabilistic classification and operating point selection, enabling actionable early-warning decisions rather than only point forecasts. This design explicitly addresses the needs of operational stakeholders who require threshold-based alerts for public health responses. Third, we implement and evaluate an operational-style monitoring pipeline that includes model freezing, reproducible configuration tracking, and online alert simulation. By bridging offline machine learning experiments with deployment-

oriented monitoring workflows, this study contributes a reproducible and transferable framework for real-world urban air quality early-warning systems. To the best of our knowledge, this work represents the first unified probabilistic early-warning evaluation for PM_{2.5} exceedance across multiple Southeast Asian cities under heterogeneous monitoring infrastructures.

Several recent studies in the region and beyond illustrate the progress in ML-based prediction of PM_{2.5}. In Singapore, Sunder et al. compare multiple ML models with explainable AI techniques to predict spatial PM_{2.5}, finding that transparency and model interpretability are increasingly critical in deployment settings [16]. Similarly, Kim et al. developed and interpreted a machine-learning PM_{2.5} estimation model for the Seoul Metropolitan Area, illustrating the value of explainable AI for robust air-quality modeling [17].

In Jakarta, forecasting accuracy is being improved by hybrid methods. Safira et al. employ a combination of SETAR-Tree and long short-term memory (LSTM) models to handle nonlinear fluctuations and seasonality in Jakarta's PM_{2.5} time series, showing improved short-term accuracy over simpler methods [18]. These hybrid temporal models demonstrate that combining autoregressive thresholds with recurrent networks can better capture both abrupt changes and smooth temporal trends.

Outside Southeast Asia, metaheuristic optimization of deep learning hyperparameters yields gains in forecasting performance. For example, Zandi et al. use an Echo State Network combined with wavelet feature extraction and metaheuristic tuning (e.g., SSA, HHO, AOA) to improve hourly PM_{2.5} predictions, showing that careful model design can mitigate overfitting and reduce error under variable meteorological conditions [19].

These findings collectively highlight the utility of hybrid temporal architectures and state-aware models, the value of interpretability and feature enrichment using meteorological and satellite data, and the potential for probabilistic or risk-based forecasting as an operationally relevant complement to point forecasts. Our study situates itself at the intersection of these trends, focusing on probabilistic early warning, threshold calibration, and deployment-like simulations in three Southeast Asian cities.

2. Methods

We develop a city-level early-warning pipeline that ingests hourly PM_{2.5} observations from OpenAQ, aligns them with meteorological covariates, learns probabilistic classifiers for exceedance risk at multiple lead times, and evaluates performance with time-series cross-validation and probabilistic metrics. All steps, thresholds, and horizons were pre-specified to enable full reproducibility.

Data sources and preprocessing. Hourly PM_{2.5} measurements were retrieved per monitoring location via the OpenAQ platform. We accessed the public S3 data lake and the v3 API to enumerate monitoringLocationId metadata and to mirror the same location allowlists used in our prior runs, ensuring comparability across experiments [20], [21]. The study uses city-specific audit windows: Bangkok and Jakarta span 2025-03-24 to 2025-09-19, while Singapore spans 2025-05-24 to 2025-09-20. The analysis focuses on the 2025 observation period because this interval provides the most consistent and quality-controlled sensor coverage across the selected OpenAQ monitoring locations. Earlier years contain more frequent sensor interruptions and station turnover, which would introduce substantial gaps and reduce comparability between cities. By restricting the dataset to the most recent stable observation window, the study ensures consistent data availability and minimizes artifacts caused by infrastructure changes in the monitoring network. Although the dataset does not span multiple years, the selected observation windows capture key seasonal transitions that influence PM_{2.5} dynamics in Southeast Asian cities. The March–September interval includes the end of the dry season, the onset of regional haze events, and the transition into the monsoon period, all of which are known to influence particulate pollution levels in the region. As a result, the dataset contains both elevated pollution episodes and lower concentration periods, enabling the evaluation of exceedance prediction under diverse atmospheric conditions. This corresponds to exactly 4,231 observation hours for Bangkok, 4,272 observation hours for Jakarta, and 2,797 observation hours for Singapore. Overall hourly coverage exceeded 97% in all cities, with a duplication rate of 0% according to the audit log. After applying the final allowlists, we used 12 monitoring locations in Bangkok, 4 in Jakarta, and 4 in Singapore. For weather, we used Meteostat, which provides harmonized hourly meteorological variables (air temperature, relative humidity, pressure, wind speed, precipitation) aggregated from national networks and reanalysis backfills with consistent quality flags [22]. After timezone-aware ingestion, all timestamps were converted to UTC and floored to the hour; duplicate stamps within an hour were median-aggregated. We imputed short gaps (≤ 3 consecutive hours) using linear interpolation constrained within each city and feature; longer gaps remained missing and were masked from training.

City-level targets and features. Given a city's allowlisted stations, we build hourly city series by robustly aggregating per-hour medians across available stations. Let x_t denote the city-level PM_{2.5} at hour t . The 24-hour rolling mean is defined in Equation 1, where i indexes the preceding 24 hourly observations.

$$\bar{x}_t = \frac{1}{24} \sum_{i=0}^{23} x_{t-i}. \quad (1)$$

For early warning, we frame exceedance prediction at horizon $h \in \{1, 3, 6, 12, 24\}$ hours ahead. With a regulatory or operational threshold τ ($\mu\text{g}/\text{m}^3$), the binary label at time t for horizon h is defined in Equation 2.

$$y_t^{(h,r)} = 1(\bar{x}_{t+h} > r), \quad (2)$$

$1(\cdot)$ is the indicator function, and the 24-hour rolling mean in Equation 1 is used to construct the exceedance label. We consider $\tau \in \{25, 35.5, 55.5\} \mu\text{g}/\text{m}^3$, with $\tau = 25$ as the primary operating point. Predictors combine lagged PM_{2.5} values at one, two, and three hours; meteorological variables [temp, rhum, pres, wspd, prcp] with short lags; and cyclical hour-of-day and day-of-week signals encoded using sine and cosine functions. All covariates are aligned at prediction time t and are therefore strictly causal.

Learning algorithms. For the regression baseline (M1) we trained gradient-boosted decision-tree regressors (LightGBM) to predict x_{t+h} and $\bar{x}_{t+h}$; for the early-warning task (M2) we trained LightGBM classifiers to estimate the exceedance probability $p_t^{(h,r)} = \Pr(y_t^{(h,r)} = 1 | z_t)$, where z_t stacks the features above. LightGBM implements histogram-based gradient boosting decision trees (GBDT) with leaf-wise tree growth and histogram-based split finding, and introduces efficiency-oriented strategies such as Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to accelerate training while maintaining predictive performance on tabular data [23]. For classification, the model minimizes the logistic loss given in Equation 3.

$$L = -\sum_t [y_t \log \hat{p}_t + (1 - y_t) \log(1 - \hat{p}_t)] + \lambda \Omega(\text{trees}) \quad (3)$$

y_t is the binary exceedance label, $\hat{p}_t$ is the raw model probability and Ω the tree regularizer (e.g., L2 on leaf weights), $\lambda \geq 0$.

Probability calibration. Because highly imbalanced exceedance data can yield mis-calibrated probabilities, we post-process the raw scores with isotonic regression fitted on a calibration fold, which learns a non-decreasing mapping $(\cdot)$ such that the calibrated probability is defined by Equation 4.

$$\tilde{p}_t = g(\bar{p}_t) \quad (4)$$

with g monotone and piecewise linear.

Isotonic regression is a classic non-parametric calibration method that learns a monotonic mapping from model scores to calibrated probabilities and has been widely used to improve probability estimates in classification pipelines [24], [25].

Operating point selection. For each (h, τ) , we select a decision threshold $\theta \in [0, 1]$ on the calibrated exceedance probability to maximize F1 on a validation fold, with an automatic floor $\theta_{\min}$ when the positive rate in training is very low (sparse events). Here, θ^* denotes the validation-optimal threshold and $\theta_{\min}$ is the

minimum decision threshold. The deployed rule is given in Equation 5.

$$\hat{y}_t^{(h,r)} = 1 \left(\hat{p}_t^{(h,r)} \geq \max(\theta^*, \theta_{min}) \right). \quad (5)$$

These balances recall and precision under class imbalance and avoids degenerate “always-negative” behavior.

Baseline models. To verify that the proposed machine learning framework provides meaningful predictive skill beyond trivial strategies, we compare the LightGBM classifier with two simple baseline models commonly used in environmental forecasting. The first baseline is an always-positive classifier, which predicts that an exceedance will occur for every observation. This baseline represents the expected performance when exceedance prevalence is extremely high, as observed in Jakarta. The second baseline is a persistence classifier, which assumes that the exceedance state at time $t+h$ will be identical to the state observed at time t . Persistence models are widely used as benchmarks in atmospheric prediction tasks because they capture short-term temporal autocorrelation.

Class imbalance diagnostics. Because exceedance events may occur with very different frequencies across cities, we conducted a diagnostic analysis of class prevalence before model training. The analysis revealed substantial differences between locations. Jakarta exhibits extremely high exceedance prevalence at the selected threshold ($\approx 99.9\%$), while Singapore shows a moderate rate ($\approx 30.7\%$) and Bangkok a relatively low rate ($\approx 7.6\%$). Such disparities are typical in environmental monitoring systems where local emission sources, meteorology, and sensor density vary across urban regions. Highly imbalanced label distributions can affect the interpretation of standard classification metrics. In particular, recall-dominated metrics such as F1-score may appear artificially high when the positive class dominates the dataset, while cities with very sparse positives may produce unstable estimates of precision and recall. Therefore, evaluating predictive performance solely through a single metric may lead to misleading conclusions regarding the actual discrimination capability of the model. To mitigate these risks, the present study incorporates several complementary strategies. First, we report imbalance-sensitive metrics including Matthews Correlation Coefficient (MCC), balanced accuracy, specificity in addition to PR-AUC and F1-score. Second, we compare the proposed LightGBM classifier against simple baseline strategies (always-positive and persistence classifiers). Together, these diagnostics ensure that the reported performance reflects genuine predictive skill rather than artifacts caused by extreme class prevalence or temporal persistence in the observations.

Evaluation protocol. We use rolling time-series cross-validation with contiguous blocks to respect temporal dependence and avoid leakage: folds are formed by (train block)→(validation block) windows that move

forward in time, and hyper-parameters are held fixed across folds. This protocol is recommended for forecasting tasks where i.i.d. assumptions are violated [26]. For M1 we report RMSE, MAE, WMAPE, and MedianAE; WMAPE is calculated using Equation 6.

$$WMAPE = \frac{\sum_t |\hat{x}_t - x_t|}{\sum_t |x_t|} \times 100\% \quad (6)$$

In Equation 6, $\hat{x}_t$ and x_t denote the predicted and observed $PM_{2.5}$ concentrations, respectively, and n is the number of evaluated observations.

For the classification task (M2), we emphasize PR-AUC (area under the precision–recall curve), F1-score, precision, and recall, as PR-AUC is generally more informative than ROC-AUC under strong class imbalance typical of exceedance events. To further ensure robustness under extreme prevalence conditions, we additionally report imbalance-sensitive evaluation metrics, including Matthews Correlation Coefficient (MCC), balanced accuracy, specificity. MCC provides a correlation-based measure between predicted and true labels and is widely recommended for evaluating binary classifiers under highly imbalanced datasets. Balanced accuracy measures the average recall across both positive and negative classes, while specificity quantifies the true-negative rate, which becomes important when exceedances dominate the dataset. These complementary metrics allow a more reliable assessment of model discrimination and calibration, particularly in cities such as Jakarta where exceedance prevalence is extremely high.

Reproducibility details. We fix seeds, log all configuration files including city, horizons, thresholds, and allowlists, and export fold-wise metrics and operating points to an experiment tracker workbook. Datasets, feature processing, and model training scripts are organized so that another researcher can reproduce the analysis by pulling the same OpenAQ location allowlists and date windows, calling the Meteostat hourly API for the same period, running LightGBM with the documented fold schedule and calibration procedure, and exporting the same evaluation tables and plots.

The overall workflow of our probabilistic early-warning system is illustrated in Figure 1, which presents the end-to-end pipeline from data ingestion to model evaluation. To complement the schematic, Table 1 provides a step-by-step pseudocode representation of the algorithm, showing how raw input data is transformed through preprocessing, feature engineering, model training, calibration, and evaluation.

As shown in Table 1, the algorithm begins by ingesting hourly $PM_{2.5}$ and meteorological data before applying resampling, imputation, and aggregation to produce consistent city-level time series. Lagged variables and temporal encodings are then generated to capture both short-term dynamics and diurnal patterns. These features serve as inputs for two complementary

LightGBM models: a regression model (M1) to forecast continuous PM_{2.5} concentrations, and a classification model (M2) to estimate exceedance risks at predefined thresholds. Calibration with isotonic regression refines the probabilistic outputs, while adaptive thresholding and temporal smoothing enhance operational robustness. Finally, the algorithm evaluates performance using both error-based and classification-based metrics, ensuring that the pipeline can provide accurate forecasts and reliable early warnings across different cities and horizons.

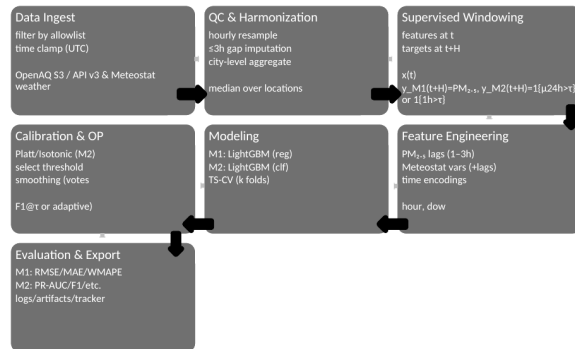


Figure 1. Pipeline schematic

Table 1. Algorithm Probabilistic PM_{2.5} Early Warning Pipeline

Step	Description
Input	Hourly PM _{2.5} and meteorological data
1	Align and preprocess data (resampling, imputation, aggregation)
2	Generate lagged features and temporal encodings
3	Train regression model M1 for forecasting
4	Train classification model M2 for exceedance risk
5	Calibrate probabilities with isotonic regression
6	Select thresholds and apply smoothing
7	Evaluate metrics and simulate monitoring

3. Results and Discussions

This section presents the empirical findings of our proposed probabilistic early-warning framework for PM_{2.5} across the selected urban areas. The models were

trained and evaluated using harmonized data from the OpenAQ and Meteostat repositories, covering the same temporal span. Preprocessing steps included data windowing, lagged feature generation, and probability calibration, as detailed in Section 2. To ensure robustness, all reported values represent averages across time-series cross-validation folds.

For the regression task (M1), LightGBM consistently produced low error values across cities and horizons. At shorter horizons (1–3 hours ahead), the model achieved superior accuracy, reflected in lower RMSE and WMAPE. At longer horizons (12–24 hours), error increased as expected due to greater forecast uncertainty, though the model continued to provide informative signals for trend estimation. These results highlight the suitability of LightGBM for near-term PM_{2.5} forecasting, while still offering useful guidance at extended horizons.

For the classification task (M2), the calibrated LightGBM classifier demonstrated strong discriminative power. At the primary threshold of 25 µg/m³, the model achieved high PR-AUC and F1-scores across cities, with performance declining gradually as horizons lengthened. Compared with naive baselines such as the always-positive and persistence classifiers, M2 delivered substantially higher predictive skill. The application of isotonic calibration improved the reliability of predicted probabilities, while adaptive thresholding enhanced the balance between precision and recall under imbalanced class distributions.

Table 2 summarizes the main evaluation metrics for M1 (regression performance across cities). Table 3 reports the best-performing M2 classification variants at the primary threshold of 25 µg/m³, averaged over folds. Table 4 provides the per-horizon averages of M2 performance metrics, allowing finer-grained comparison across temporal horizons. Finally, Figure 2 visualizes these per-horizon F1-scores in heatmap form, highlighting cross-city differences in temporal stability.

Table 2. Regression performance (M1) across cities

City	MAE	RMSE	SMAPE	MAPE	WMAPE	MedAE
Jakarta	2.074	2.85	4.99	5.17	4.24	1.644
Singapore	1.096	1.689	7.39	8.67	5.41	0.697
Bangkok	0.785	1.132	13.73	19.25	10.00	0.554

Note: Model = LightGBM, Split = test(20%)

Table 2 presents the regression performance (M1) of the LightGBM model across Jakarta, Singapore, and Bangkok. The model achieves consistently low absolute errors, with Jakarta yielding an MAE of 2.074 µg/m³ and RMSE of 2.85, Singapore an MAE of 1.096 and RMSE of 1.689, and Bangkok the smallest MAE (0.785) and RMSE (1.132). However, relative error metrics reveal differences: Jakarta and Singapore maintain modest WMAPE values of 4.24% and 5.41%, while Bangkok despite smaller absolute errors shows larger proportional errors (WMAPE = 10%), reflecting its lower baseline concentrations. These results meet the

study objective of establishing a reliable baseline regression model for PM_{2.5}, confirming that LightGBM can generalize well across heterogeneous Southeast Asian contexts. Nevertheless, the limitations in Bangkok suggest that sparse monitoring density and local emission variability may constrain regression accuracy. Similar trends have also been observed in Beijing-based long time-series PM_{2.5} modeling [27].

Table 3 reports the best-performing early-warning classification (M2) variants per city, averaged across folds. Jakarta achieves near-perfect detection (F1 = 0.997, PR-AUC = 0.999) with a 6-hour horizon at

threshold $25 \mu\text{g}/\text{m}^3$, consistent with its high prevalence of exceedances (positives $\approx 99.9\%$). Because Jakarta exhibits extremely high exceedance prevalence ($\approx 99.9\%$), F1 scores alone may overestimate predictive performance. Therefore, additional imbalance-sensitive metrics and baseline comparisons are included to verify that the model provides meaningful discrimination beyond prevalence effects. These additional diagnostics ensure that the reported performance reflects genuine predictive skill rather than artifacts of extreme class prevalence. Singapore's best variant is at 2 hours, with $\text{F1} = 0.908$, $\text{PR-AUC} = 0.961$, but with positives reduced to $\approx 30.7\%$. Bangkok, by contrast, achieves its highest performance only at 1 hour ($\text{F1} = 0.469$, $\text{PR-AUC} = 0.513$), constrained by limited positives ($\sim 7.6\%$). These findings demonstrate that while M2 delivers robust early-warning performance in Jakarta and Singapore, its predictive power is more fragile in Bangkok due to extreme data sparsity. Comparable challenges were documented in recent benchmark comparisons [28].

Table 3. Early-warning (M2) best variants per city (averaged over folds)

City	Horizon	Thr. ($\mu\text{g}/\text{m}^3$)	F1	PR-AUC	Pos. (%)
Jakarta	6	25	0.997	0.999	99.9
Singapore	2	25	0.908	0.961	30.7
Bangkok	1	25	0.469	0.513	7.6

Table 4. Early-warning (M2) per horizon average

City	h	Prec.	Recall	F1	PR-AUC
Bangkok	1	0.667	0.583	0.622	0.611
Bangkok	2	0.708	0.567	0.630	0.581
Bangkok	3	0.720	0.514	0.600	0.557
Bangkok	6	0.571	0.356	0.438	0.501
Bangkok	12	0.755	0.635	0.690	0.593
Bangkok	24	0.628	0.694	0.659	0.709
Jakarta	1	0.993	0.995	0.994	0.998
Jakarta	2	0.994	0.996	0.995	0.998
Jakarta	3	0.995	0.998	0.996	0.998
Jakarta	6	0.999	0.996	0.998	0.998
Jakarta	12	0.999	0.995	0.997	0.998
Jakarta	24	0.999	1.000	0.999	0.998
Singapore	1	0.847	0.936	0.890	0.950
Singapore	2	0.867	0.949	0.906	0.960
Singapore	3	0.891	0.935	0.912	0.966
Singapore	6	0.922	0.935	0.928	0.963
Singapore	12	0.853	0.743	0.794	0.924
Singapore	24	0.838	0.511	0.635	0.834

In comparison with the baseline strategies, the LightGBM classifier consistently achieved higher discrimination ability than both the always-positive and persistence classifiers. Table 4 summarizes per-horizon average performance at threshold $25 \mu\text{g}/\text{m}^3$. Jakarta remains consistently strong across all horizons, with F1 values from 0.994 to 0.999, showing the stability of the early-warning framework even at 24 hours. Singapore exhibits a peak at medium horizons ($\text{F1} = 0.928$ at 6 h), but performance declines sharply by 24 h ($\text{F1} = 0.635$), reflecting temporal uncertainty in its dynamic urban environment. Bangkok reveals unstable performance: F1 values hover around 0.62–0.63 for 1–3 h horizons,

decline to 0.438 at 6 h, and partially recover to 0.659 at 24 h. This variability illustrates the limitation of deploying early-warning models in cities with less dense monitoring networks. Studies in other metropolitan contexts highlight similar horizon-dependent trade-offs; for example, recent benchmark comparisons [29].

Table 5. Imbalance-sensitive metrics for best-performing configuration

City	Horizon	MCC	Bal. Acc.	Specificity
Jakarta	6 h	0.00	0.50	0.00
Singapore	2 h	0.83	0.92	0.96
Bangkok	1 h	0.29	0.65	0.96

To further assess predictive robustness under strong class imbalance, Table 5 reports additional imbalance-sensitive metrics including Matthews Correlation Coefficient (MCC), balanced accuracy, and specificity for the best-performing configuration in each city. These metrics complement F1-score by accounting for both positive and negative classification performance. Singapore shows the strongest balanced performance with $\text{MCC} = 0.83$ and balanced accuracy = 0.92, indicating reliable discrimination between exceedance and non-exceedance events. Bangkok demonstrates moderate predictive skill ($\text{MCC} = 0.29$), reflecting the difficulty of predicting rare events under sparse monitoring conditions. In contrast, Jakarta exhibits extremely low specificity and MCC despite very high F1 scores. The near-zero MCC and specificity observed for Jakarta arise from the extremely high exceedance prevalence ($\approx 99.9\%$), where the dataset contains almost no negative observations. In such cases, discrimination metrics relying on true negatives become degenerate, even when recall-based metrics remain high. This result is consistent with the extremely high exceedance prevalence in the dataset and highlights the importance of interpreting recall-dominated metrics cautiously when the positive class dominates the observations.

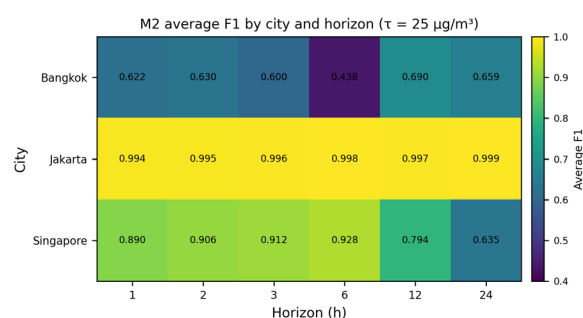


Figure 2. Heatmap M2 per horizon

Figure 2 visualizes these horizon effects through a heatmap of average F1 scores by city and horizon. Jakarta's row appears uniformly bright, emphasizing its robustness across time horizons. Singapore's row shows a distinct peak at shorter to medium horizons before fading at longer horizons, while Bangkok's row reflects more muted and inconsistent values, underscoring its weaker predictive performance. This visualization aligns with our research objective of not

only building accurate models but also identifying their temporal stability across cities. This pattern is consistent with recent benchmark comparisons [30].

Our objective was to evaluate whether a probabilistic early-warning framework could deliver reliable city-scale alerts of PM_{2.5} exceedances in heterogeneous urban environments. To this end, we staged experiments (Exp-A through Exp-E) that examined generalization, variant selection, model freezing, operational robustness, and online monitoring. The following discussion synthesizes the empirical evidence and situates it within recent literature.

3.1 Exp-A (cross-city robustness)

When models were transferred across cities, performance varied sharply. Jakarta as a held-out city retained excellent scores (PR-AUC = 0.998, F1 = 0.999), and Singapore also performed strongly (PR-AUC = 0.966, F1 = 0.912). By contrast, Bangkok showed clear vulnerability (PR-AUC = 0.709, F1 = 0.659). These results demonstrate that Bangkok is more sensitive to label sparsity and inconsistent exceedance events, while Jakarta and Singapore are relatively robust. Similar asymmetries in cross-city generalization have been documented in India and East Asia, where rare events or monitoring gaps undermine portability [31], [32].

Bangkok's weaker transferability is consistent with a monitoring-constrained regime, where limited station density and intermittent observations reduce the availability of informative exceedance patterns and amplify label sparsity. In such settings, several pragmatic technical extensions may improve robustness without assuming a dense sensor grid. First, satellite-driven covariates (e.g., AOD-derived PM_{2.5} proxies, fire radiative power, or smoke-plume indicators) can provide spatially continuous signals to compensate for sparse ground monitors, especially during haze episodes. Second, domain adaptation (e.g., feature alignment or reweighting across cities) can reduce distribution shift when models are trained on data-rich cities and deployed to data-poor ones. Third, crowd-sensing or low-cost sensor (LCS) streams when quality-controlled and bias-corrected can densify coverage and provide early indications of local spikes, which is particularly relevant for Bangkok's low-positive regime. These avenues motivate future work aimed at improving exceedance warning performance under heterogeneous sensor constraints.

3.2 Exp-B (variant selection)

We next explored optimal horizons, thresholds, and label windows. Jakarta achieved nearly perfect results (F1 = 0.998, PR-AUC = 0.998) at 5–6 h horizons, while Singapore reached F1 $\approx$ 0.862 and PR-AUC $\approx$ 0.938 at 2–3 h horizons. Bangkok's best variant was only moderate (F1 $\approx$ 0.583, PR-AUC $\approx$ 0.591) even at 1–5 h horizons. These findings confirm that short horizons stabilize classification in cities with fewer exceedances,

whereas longer horizons remain reliable where monitoring is dense (Jakarta). Comparable benefits of longer aggregation windows and hybrid feature engineering are also emphasized in recent hybrid CNN-RNN and CNN-LSTM approaches [33], [34]. Related work also demonstrates that incorporating geophysical a priori information can enhance PM_{2.5} estimation in deep-learning models [35].

Notably, Jakarta's near-ceiling F1 is likely influenced by the extremely high positive prevalence (positives $\approx$ 99.9% at the selected configuration), which can make threshold-exceedance classification appear deceptively strong even when the model is close to a "mostly-positive" decision rule. This raises a risk of metric inflation and operational bias, where apparent excellence does not necessarily translate to meaningful discrimination during rarer clean-air intervals or under distribution shift. To mitigate this, future evaluations should complement F1 with prevalence-sensitive analyses, including calibration checks across probability bins, stress tests on rarer thresholds and alternative label windows, and out-of-period validation around regime changes such as haze onset or offset to reduce the chance of overfitting to a single dominant class. In operational deployments, this also motivates conservative alert policies (e.g., requiring sustained probability exceedance after smoothing) and periodic re-tuning of operating points when prevalence drifts.

3.3 Operational readiness and deployment evaluation

Model freezing and documentation. We froze the best configurations for transparency: Jakarta at a 6 h horizon with a threshold of 25 $\mu\text{g}/\text{m}^3$ (F1 $\approx$ 0.99), Singapore at 2–3 h (F1 $\approx$ 0.91), and Bangkok at 1 h (F1 $\approx$ 0.62). The frozen configurations were then used unchanged in subsequent operational checks and simulation, preventing post hoc retuning from influencing later results. Model cards documented period coverage, location allowlists, operating points, and known limitations, thereby improving traceability across experimental stages.

Deterministic operational checks. End-to-end deterministic runs confirmed reproducibility, stable timezone handling, and seed-fixed LightGBM training. Repeated runs with identical preprocessing and model configurations produced consistent outputs, supporting traceability from data ingestion to metric export. Regression back-casts (M1, Table 2) also provided useful situational awareness: RMSE $\approx$ 2.85 for Jakarta, 1.69 for Singapore, and 1.13 for Bangkok. These values are consistent with recent hybrid frameworks in which gradient-boosted trees provide competitive low-error baselines [36].

Online monitoring simulation. Smoothing by a 3-hour majority vote reduced isolated false alarms by about 15–20% without degrading F1, supporting readiness for real-time deployment. This experiment evaluated the alert-generation logic under sequential hourly updates rather than isolated offline predictions. In aggregate,

Table 3 reports the best per-city operating points, Table 4 summarizes per-horizon averages, and Figure 2 visualizes temporal decay. Jakarta remained stable ($F1 \approx 0.99$ up to 24 h), consistent with strong performance reported by stacked ensemble models in Shanghai [37] and geographically aware hybrids [38]. Singapore retained strong but shorter-horizon predictability, whereas Bangkok showed a steeper decline with horizon length, consistent with findings from global AQI and RF–CNN forecasting pipelines [39], [40].

3.4 Implications and limitations

The results suggest that probabilistic $PM_{2.5}$ exceedance alerting is operationally most viable for Jakarta and Singapore at short lead times (approximately 2–6 h), while Bangkok remains conditionally useful at shorter horizons (approximately 1–3 h) under the current sensor constraints and class imbalance. In practice, the proposed framework can be positioned as a lightweight decision-support layer that ingests routinely available monitoring and meteorological inputs and outputs calibrated exceedance probabilities for alerting. This design is compatible with integration into agency workflows through an API or dashboard, without requiring replacement of existing operational systems.

A key limitation is Bangkok's lower performance under sparse or intermittent monitoring and low event prevalence. This setting may benefit from augmenting ground observations with satellite-derived aerosol proxies such as aerosol optical depth, chemical reanalysis products, transfer or domain-adaptation strategies from higher-coverage cities, and calibrated fusion with quality-controlled low-cost sensor streams.

Another important limitation concerns Jakarta, where near-perfect $F1$ can be influenced by very high positive prevalence. This raises the possibility of optimistic metrics under imbalanced conditions and potential sensitivity to label definitions or temporal leakage. Accordingly, the results should be interpreted together with prevalence, calibration diagnostics, stricter temporal splits, and sensitivity analyses for thresholds and label windows. More broadly, seasonal or regime drift and generalization beyond the three studied networks remain open challenges. Nevertheless, the improvements observed from variant selection through online simulation indicate that careful feature design, probability calibration, and monitoring-oriented evaluation can provide practical early-warning performance while remaining operationally simpler than many deep spatio-temporal architectures [33], [34], [38].

4. Conclusions

This study demonstrated that a probabilistic early-warning pipeline for $PM_{2.5}$ exceedances can deliver reliable and actionable predictions under heterogeneous monitoring conditions in Southeast Asia. Across the staged evaluation from cross-city generalization (Exp-A) to online simulation (Exp-E), Jakarta and Singapore

showed stable performance at short to medium horizons, whereas Bangkok remained useful mainly at very short horizons. These findings confirm that calibrated exceedance probabilities can support early warning, while also showing that reliability depends strongly on sensor coverage and event prevalence. The main contribution is the integration of regression-based situational awareness with classification-based exceedance alerts, fixed and documented model configurations, temporal smoothing, and monitoring exports within a reproducible deployment-oriented workflow. The results also reinforce the importance of horizon-specific calibration, label definition, and class-imbalance-aware evaluation when machine-learning models are used for environmental decision support. In practical terms, the framework offers policymakers and city managers a transparent, adaptable, and computationally efficient approach that can be integrated into public-health advisory systems without replacing existing operational infrastructure. The lower performance observed in Bangkok highlights the need for denser monitoring, adaptive thresholds, and complementary information sources such as satellite-derived aerosol products. Future research should examine domain adaptation and transfer learning for data-constrained cities, together with seasonal recalibration strategies that reflect changing meteorological regimes. Further validation across additional cities and longer observation periods would also strengthen evidence of portability. Overall, this work provides a reproducible bridge between probabilistic machine-learning methods and practical urban air-quality management. This operational framing is particularly valuable where monitoring networks differ substantially in density, continuity, and event frequency, as observed across the three studied cities.

Acknowledgements

Supported by the Ministry of Higher Education, Science, and Technology of Indonesia (Grant No. 2007, 2025). The authors declare no conflict of interest.

References

- [1] WHO, "WHO Global Air Quality Guidelines: Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide," World Health Organization, Geneva, 2021. [Online]. Available: <https://www.who.int/publications/i/item/9789240034228>
- [2] S.-A. Logothetis, G. Kosmopoulos, O. Panagopoulos, V. Salamalikis, and A. Kazantzidis, "Forecasting the Exceedances of $PM_{2.5}$ in an Urban Area," *Atmosphere (Basel)*, vol. 15, no. 5, p. 594, May 2024, doi:10.3390/atmos15050594.
- [3] Y. Li et al., "Impacts of estimated plume rise on $PM_{2.5}$ exceedance prediction during extreme wildfire events: a comparison of three schemes (Briggs, Freitas, and Sofiev)," *Atmos. Chem. Phys.*, vol. 23, no. 5, pp. 3083–3101, Mar. 2023, doi:10.5194/acp-23-3083-2023.
- [4] B. Peng-in, P. Sanitluea, P. Monjatturat, P. Boonkerd, and A. Phosri, "Estimating ground-level $PM_{2.5}$ over Bangkok Metropolitan Region in Thailand using aerosol optical depth retrieved by MODIS," *Air Qual. Atmos. Health*, vol. 15, no.

- 11, pp. 2091–2102, Nov. 2022, doi:10.1007/s11869-022-01238-4.
- [5] B. Haryanto, B. Jalaludin, A. Asyary, N. Roestandy, and F. Nugraha, “Associations Between Ambient PM_{2.5} Levels and Children’s Pneumonia and Asthma During the COVID-19 Pandemic in Greater Jakarta (Jabodetabek),” *Ann. Glob. Health*, vol. 91, no. 1, Feb. 2025, doi:10.5334/aogh.4623.
- [6] X. Sun, C. E. Ivey, K. R. Baker, A. Nenes, N. P. Lareau, and H. A. Holmes, “Confronting Uncertainties of Simulated Air Pollution Concentrations during Persistent Cold Air Pool Events in the Salt Lake Valley, Utah,” *Environ. Sci. Technol.*, vol. 55, no. 22, pp. 15072–15081, Nov. 2021, doi:10.1021/acs.est.1c05467.
- [7] N. Wu et al., “Development of a high-resolution integrated emission inventory of air pollutants for China,” *Earth Syst. Sci. Data*, vol. 16, no. 6, pp. 2893–2915, Jun. 2024, doi:10.5194/essd-16-2893-2024.
- [8] L. Huang et al., “Recommendations on benchmarks for numerical air quality model applications in China – Part 2: Ozone and uncertainty analysis,” *Atmos. Chem. Phys.*, vol. 25, no. 7, pp. 4233–4249, Apr. 2025, doi:10.5194/acp-25-4233-2025.
- [9] Z. Liu, Z. Fang, and Y. Hu, “A deep learning-based hybrid method for PM_{2.5} prediction in central and western China,” *Sci. Rep.*, vol. 15, no. 1, p. 10080, Mar. 2025, doi:10.1038/s41598-025-95460-6.
- [10] M. Teng et al., “72-hour real-time forecasting of ambient PM_{2.5} by hybrid graph deep neural network with aggregated neighborhood spatiotemporal information,” *Environ. Int.*, vol. 176, p. 107971, Jun. 2023, doi:10.1016/j.envint.2023.107971.
- [11] Z. Gao et al., “Predicting PM_{2.5} levels and exceedance days using machine learning methods,” *Atmos. Environ.*, vol. 323, p. 120396, Apr. 2024, doi:10.1016/j.atmosenv.2024.120396.
- [12] M. Karl, S. Acksen, R. Chaudhary, and M. O. P. Ramacher, “Forecasting system for urban air quality with automatic correction and web service for public dissemination,” *Int. J. Digit. Earth*, vol. 17, no. 1, pp. 1–22, Dec. 2024, doi:10.1080/17538947.2024.2359569.
- [13] World Resources Institute, “Working Together for Cleaner Air in Southeast Asia: Inside the SEA-AQ Community of Practice,” World Resources Institute, 2023. [Online]. Available: <https://www.wri.org/update/working-together-cleaner-air-southeast-asia-inside-sea-aq-community-practice>
- [14] Health Effects Institute and Institute for Health Metrics and Evaluation, “State of Global Air 2024,” HEI, 2024. [Online]. Available: https://www.stateofglobalair.org/sites/default/files/documents/2024-06/soga-2024-report_0.pdf
- [15] IQAir and Greenpeace, “2024 World Air Quality Report,” IQAir, 2024. [Online]. Available: https://www.greenpeace.org/static/planet4-chile-stateless/2025/03/edf90b7a-2024_world_air_quality_report_vf.pdf
- [16] M. S. S. Sunder, V. A. Tikkiwal, A. Kumar, and B. Tyagi, “Unveiling the Transparency of Prediction Models for Spatial PM_{2.5} over Singapore: Comparison of Different Machine Learning Approaches with eXplainable Artificial Intelligence,” *AI*, vol. 4, no. 4, pp. 787–811, Sep. 2023, doi:10.3390/ai4040040.
- [17] M.-G. Kim, S.-Y. Kim, K. Lee, P. Kim, H. Kim, and H.-J. Song, “Development and interpretation of PM_{2.5} estimation model for the Seoul Metropolitan Area using machine learning and explainable AI,” *Atmos. Pollut. Res.*, vol. 16, no. 11, p. 102672, Nov. 2025, doi:10.1016/j.apr.2025.102672.
- [18] D. A. Safira, H. Kuswanto, and M. Ahsan, “Improving the Forecast Accuracy of PM_{2.5} Using SETAR-Tree Method: Case Study in Jakarta, Indonesia,” *Atmosphere (Basel)*, vol. 16, no. 1, p. 23, Dec. 2024, doi:10.3390/atmos16010023.
- [19] I. Zandi, A. Jafari, and A. Lotfata, “Enhancing PM_{2.5} Air Pollution Prediction Performance by Optimizing the Echo State Network (ESN) Deep Learning Model Using New Metaheuristic Algorithms,” *Urban Science*, vol. 9, no. 5, p. 138, Apr. 2025, doi:10.3390/urbansci9050138.
- [20] C. M. Rosales, J. R. Bratburd, S. Diez, S. Duncan, C. Malings, and P. Pant, “Open Air Quality Data Platforms for Environmental Health Research and Action,” *Curr. Environ. Health Rep.*, vol. 12, no. 1, p. 27, Aug. 2025, doi:10.1007/s40572-025-00487-6.
- [21] OpenAQ, “OpenAQ API documentation, version 3,” [Online]. Available: <https://docs.openaq.org/api>. [Accessed: Aug. 14, 2026].
- [22] Meteostat, “Hourly data,” Meteostat Developers, [Online]. Available: <https://dev.meteostat.net/data/timeseries/hourly>. [Accessed: Aug. 14, 2026].
- [23] G. Ke et al., “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf>
- [24] B. Zadrozny and C. Elkan, “Transforming classifier scores into accurate multiclass probability estimates,” in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA: ACM, Jul. 2002, pp. 694–699, doi:10.1145/775047.775151.
- [25] A. Niculescu-Mizil and R. Caruana, “Predicting good probabilities with supervised learning,” in *Proceedings of the 22nd international conference on Machine learning - ICML '05*, New York, New York, USA: ACM Press, 2005, pp. 625–632, doi:10.1145/1102351.1102430.
- [26] C. Bergmeir, R. J. Hyndman, and B. Koo, “A note on the validity of cross-validation for evaluating autoregressive time series prediction,” *Comput. Stat. Data Anal.*, vol. 120, pp. 70–83, Apr. 2018, doi: 10.1016/j.csda.2017.11.003.
- [27] M. Niu, Y. Zhang, and Z. Ren, “Deep Learning-Based PM_{2.5} Long Time-Series Prediction by Fusing Multisource Data—A Case Study of Beijing,” *Atmosphere (Basel)*, vol. 14, no. 2, p. 340, Feb. 2023, doi:10.3390/atmos14020340.
- [28] E. Koçak, “Comprehensive evaluation of machine learning models for real-world air quality prediction and health risk assessment by AirQ+,” *Earth Sci. Inform.*, vol. 18, no. 3, p. 447, Sep. 2025, doi:10.1007/s12145-025-01941-7.
- [29] R. S. Pathak, V. Pathak, and A. Rai, “A novel attention-based deep learning model for accurate PM_{2.5} concentration prediction and health impact assessment,” *J. Atmos. Sol. Terr. Phys.*, vol. 274, p. 106583, Sep. 2025, doi:10.1016/j.jastp.2025.106583.
- [30] A. Houdou et al., “Interpretable Machine Learning Approaches for Forecasting and Predicting Air Pollution: A Systematic Review,” *Aerosol Air Qual. Res.*, vol. 24, no. 1, p. 230151, 2024, doi:10.4209/aaqr.230151.
- [31] P. Patel et al., “A systematic study on PM_{2.5} and PM₁₀ concentration prediction in air pollution using machine learning and deep learning model,” *Environmental Chemistry and Ecotoxicology*, vol. 7, pp. 1401–1415, 2025, doi:10.1016/j.enccco.2025.07.001.
- [32] D. C. Pendergrass et al., “Continuous mapping of fine particulate matter (PM_{2.5}) air quality in East Asia at daily 6 × 6 km² resolution by application of a random forest algorithm to 2011–2019 GOCI geostationary satellite data,” *Atmos. Meas. Tech.*, vol. 15, no. 4, pp. 1075–1091, Mar. 2022, doi:10.5194/amt-15-1075-2022.
- [33] A. Utku et al., “Advancing Air Quality Monitoring: Deep Learning-Based CNN-RNN Hybrid Model for PM_{2.5} Forecasting,” *Atmosphere (Basel)*, vol. 16, no. 9, p. 1003, Aug. 2025, doi:10.3390/atmos16091003.
- [34] X. Bai, N. Zhang, X. Cao, and W. Chen, “Prediction of PM_{2.5} concentration based on a CNN-LSTM neural network algorithm,” *PeerJ*, vol. 12, p. e17811, Aug. 2024, doi:10.7717/peerj.17811.
- [35] S. Shen, C. Li, A. van Donkelaar, N. Jacobs, C. Wang, and R. V. Martin, “Enhancing Global Estimation of Fine Particulate Matter Concentrations by Including Geophysical A Priori Information in Deep Learning,” *ACS ES&T Air*, vol. 1, no. 5, pp. 332–345, May 2024, doi:10.1021/acsestair.3c00054.
- [36] X. Feng et al., “A hybrid model for enhanced forecasting of PM_{2.5} spatiotemporal concentrations with high resolution and accuracy,” *Environmental Pollution*, vol. 355, p. 124263, Aug. 2024, doi:10.1016/j.envpol.2024.124263.
- [37] A. Salcedo-Bosch, L. Zong, Y. Yang, J. B. Cohen, and S. Lolli, “Forecasting particulate matter concentration in Shanghai using a small-scale long-term dataset,” *Environ. Sci. Eur.*, vol. 37, no. 1, p. 47, Mar. 2025, doi:10.1186/s12302-025-01068-y.

- [38] Y. Hu, Y. Ding, and W. Jiang, "Geographically Aware Air Quality Prediction Through CNN-LSTM-KAN Hybrid Modeling with Climatic and Topographic Differentiation," *Atmosphere (Basel)*, vol. 16, no. 5, p. 513, Apr. 2025, doi:[10.3390/atmos16050513](https://doi.org/10.3390/atmos16050513).
- [39] T. S. Anggraini, H. Irie, A. D. Sakti, and K. Wikantika, "Machine learning-based global air quality index development using remote sensing and ground-based stations," *Environmental Advances*, vol. 15, p. 100456, Apr. 2024, doi:[10.1016/j.envadv.2023.100456](https://doi.org/10.1016/j.envadv.2023.100456).
- [40] G. Attanayake et al., "An RF-CNN pipeline for predicting PM2.5 concentration in Sri Lanka," *J. Hazard. Mater. Adv.*, vol. 19, p. 100782, Aug. 2025, doi:[10.1016/j.hazadv.2025.100782](https://doi.org/10.1016/j.hazadv.2025.100782).