THE ENGINEER STORY

e-ISSN: 3009 - 0792 Volume 23, 2026, 4-7

FORECASTING $PM_{2.5}$ AT FOUR MALAYSIAN STATIONS WITH DEEP LEARNING

Omer Adnan Alawi * Haslinda Mohamed Kamar
Department of Thermofluids, Faculty of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.
* Corresponding author: omeralawi@utm.my

ABSTRACT

A temporal convolutional neural network (CNN) and gradient boosting were combined to forecast $PM_{2.5}$ at four Malaysian stations. Across 17,243 test forecasts from July to December 2020, the blend achieved RMSE 4.847 and R^2 0.586 for the next six-hour mean. Its RMSE was 39.1% lower than persistence and 1.9% lower than boosting alone. Below the station threshold, average precision for a high reading within six hours rose from 0.534 to 0.553 compared with boosting. The blend missed the largest spike.

KEYWORD

$PM_{2.5}$ forecasting, temporal convolutional neural networks, gradient boosting, high-concentration alerts, Malaysia

INTRODUCTION

Short-term forecasts of fine particulate matter ($PM_{2.5}$) can give air-quality managers time to respond to rising concentrations. Yohanna and Lim (2026) found regional and seasonal differences in daily $PM_{2.5}$ across five regions of Malaysia. In southern Thailand, Damkliang and Chumnaul (2025) tested a deep learning model for daily $PM_{2.5}$ forecasts at four monitoring stations near Hat Yai.

For hourly Malaysian data, Ismael et al. (2026) combined a stacked autoencoder with convolutional and recurrent layers to predict $PM_{2.5}$. Across four Chinese cities, Cheng et al. (2026) reported lower six-hour RMSE for a multiresolution network than for a gated recurrent unit. Paraschiv et al. (2026) found that persistence outperformed random forest and XGBoost for one-hour $PM_{2.5}$ forecasts. These findings motivate comparing models across lead times and retaining persistence as a baseline.

This study tests whether blending a residual temporal convolutional neural network with histogram gradient boosting improves forecasts at Seberang Jaya, Larkin, Keningau, and Kuching. Three chronological test periods are used to evaluate hourly forecasts up to 24 hours ahead, and forecasts of the next six-hour mean. The study also estimates the probability of reaching the station's 90th percentile within six hours when the latest reading is below that threshold.

MATERIAL AND METHODOLOGY

The study used 2020 hourly pollutant and weather records from four Malaysian stations (Table 1). We forward-filled input gaps for up to three hours and then filled them with training medians. A three-layer residual temporal CNN used 24-hour histories to predict $PM_{2.5}$ at 1, 3, 6, 12, and 24 hours, the next six-hour mean, and the probability of at least one reading reaching the station's January–May 90th percentile within six hours. Validation data set the gradient boosting blend weights. Three expanding-window tests used RMSE, MAE, and R^2 for concentrations and average precision (AP), Brier score, precision, and recall for alerts.

Table 1: $PM_{2.5}$ coverage and high-reading thresholds by station.

Station	$PM_{2.5}$ hours	Test cases	Threshold
Seberang Jaya, Pulau Pinang (CA06)	8,754	4,326	30.55
Larkin, Johor (CA33)	8,777	4,322	21.90
Keningau, Sabah (CA52)	8,750	4,299	19.02
Kuching, Sarawak (CA65)	8,767	4,296	14.63

The workbooks omit concentration units, so results use the original scale.

RESULT AND DISCUSSION

Concentration Forecasts

The CNN and boosting blend had the lowest pooled RMSE at four of the five hourly leads. The CNN using $PM_{2.5}$ and weather had the lowest 24-hour RMSE (8.447). At six hours, the CNN variants with and without co-pollutants gave similar RMSE (8.148 and 8.152).

Table 2: Pooled RMSE at each lead and for the next six-hour mean.

Model	**1 h**	**3 h**	**6 h**	**12 h**	**24 h**	**6 h mean**
Persistence	8.02	9.99	11.04	11.38	10.76	7.96
Daily seasonal baseline	10.57	10.56	10.54	10.75	10.76	7.13
Ridge regression	6.70	7.75	8.22	8.45	8.50	5.00
Gradient boosting	6.57	7.54	8.15	8.50	8.57	4.94
CNN with $PM_{2.5}$ and weather	6.75	7.66	8.15	8.39	**8.45**	4.95
CNN with co-pollutants	6.69	7.70	8.15	8.41	8.49	4.94
CNN and boosting blend	**6.53**	**7.52**	**8.07**	**8.36**	8.46	**4.85**

The blend achieved MAE 3.299 and R^2 0.586 for the next six-hour mean. Its RMSE was lower than boosting across stations and test periods (Figure 1 and Table 3). The paired RMSE difference was −0.096 (95% interval −0.138 to −0.053). The interval excluded zero for 3-, 7-, and 14-day blocks.

Table 3: Chronological splits and next six-hour mean RMSE in 2020.

Training	**Validation**	**Test (forecasts)**	**Ridge**	**Boosting**	**Blend**
Jan–May	Jun	Jul–Aug (5,760)	5.622	5.569	**5.447**
Jan–Jul	Aug	Sep–Oct (5,742)	5.192	5.124	**5.049**
Jan–Sep	Oct	Nov–Dec (5,741)	4.069	3.999	**3.911**

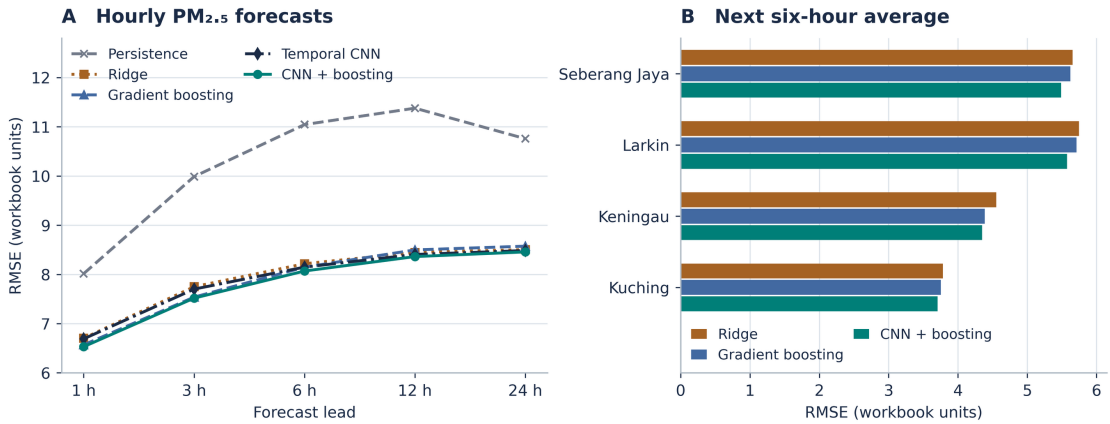


Figure 1: Pooled RMSE by lead time and next six-hour mean RMSE by station.

High-Concentration Alerts

Of 15,123 forecasts issued below the station threshold, 3,849 (25.5%) preceded at least one high reading within six hours.

Table 4: Six-hour alert performance overall and below the station threshold.

Model	All AP	Below AP	Brier	Precision	Recall
Latest $PM_{2.5}$ only	0.615	0.403	0.176	37.3%	57.4%
Logistic regression	0.679	0.521	0.158	46.3%	64.5%
Gradient boosting	0.682	0.534	0.156	50.3%	58.5%
CNN with $PM_{2.5}$ and weather	0.689	0.543	0.155	48.3%	60.6%
CNN with co-pollutants	0.687	0.534	0.156	46.5%	66.4%
CNN and boosting blend	**0.699**	**0.553**	**0.153**	49.0%	60.6%

Brier score, precision, and recall are for below-threshold forecasts. Alert triggers maximize validation F1.

The blend's below-threshold AP was 0.019 higher than boosting (95% interval 0.008 to 0.029). The blend detected 60.6% of high windows with 49.0% precision. Boosting achieved 58.5% recall and 50.3% precision. The paired bootstrap used 2,500 seven-day block resamples, with matching dates across stations and models within each test period.

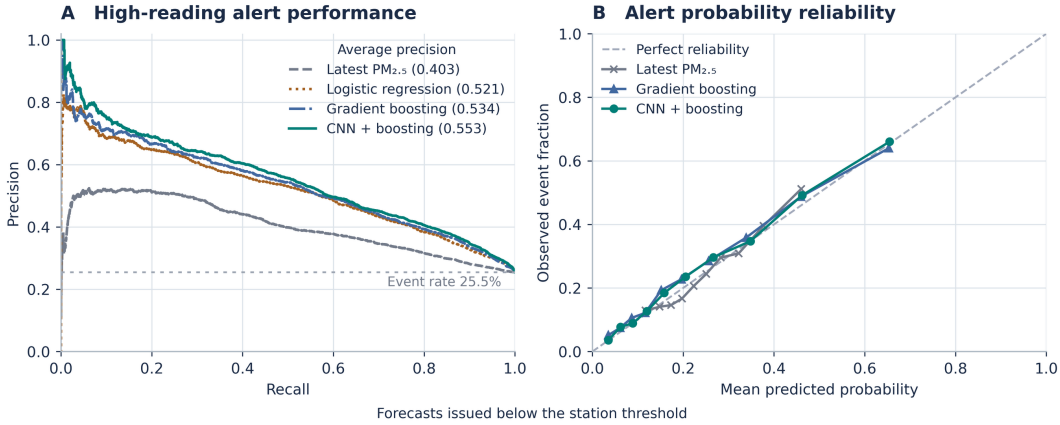


Figure 2: Alert precision–recall and reliability below the station threshold.

Performance During High-$PM_{2.5}$ Periods

The blend had higher AP than boosting in all three test periods. The CNN using $PM_{2.5}$ and weather had the highest below-threshold AP in September–October (0.592).

For the 2,134 six-hour targets at or above the station threshold, the blend underestimated $PM_{2.5}$ by 12.43 units on average. Six hours before the Keningau peak of 278.4, it forecast 7.1. Alert probability peaked at 6.4% in the preceding six hours. The trigger was 31.0%.

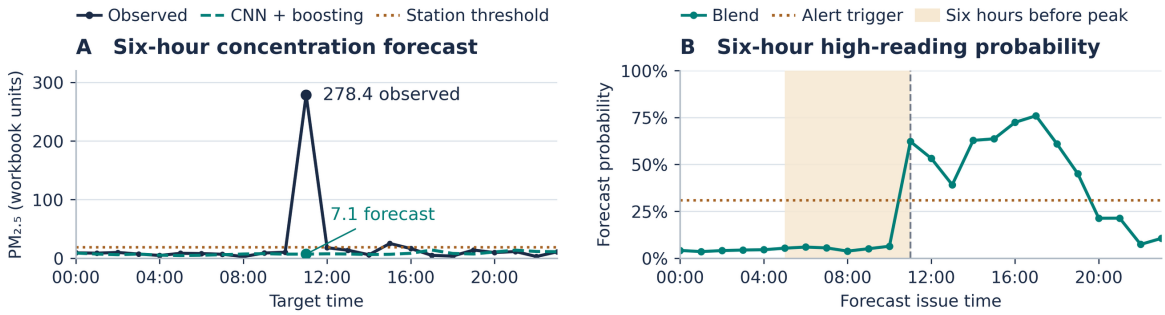


Figure 3: Keningau on 30 August 2020 by target time (A) and issue time (B).

CONCLUSION

Blending a temporal CNN with gradient boosting modestly improved the next six-hour mean forecast and alert ranking. The blend missed the largest spike. Tests covered six months at stations represented in training, and adjacent alert windows may contain the same high reading.

ACKNOWLEDGEMENT

The author thanks the Department of Environment Malaysia for making air quality monitoring information publicly available.

REFERENCE

Cheng, S., Wen, X., Xiao, J., Fan, W., & Zhou, C. (2026). Hourly PM2.5 forecasting for sustainable urban air quality management through multiresolution learning and environmental context. Sustainability, 18(20), Article 10531. https://doi.org/10.3390/su182010531

Damkliang, K., & Chumnaul, J. (2025). Deep learning and statistical approaches for area-based PM2.5 forecasting in Hat Yai, Thailand. Journal of Big Data, 12, Article 36. https://doi.org/10.1186/s40537-025-01079-9

Ismael, D. S., Masseran, N., & Abu Bakar, S. (2026). Air pollution prediction based on stacked deep autoencoder network model. Electronics, 15(13), Article 2756. https://doi.org/10.3390/electronics15132756

Paraschiv, L. S., Paraschiv, S., & Buruiana, D. L. (2026). Monitoring and forecasting of urban air pollution and exceedance events: Implications for environmental risk assessment and early warning communication using machine learning in sustainable cities. Sustainable Cities and Society, 151, Article 108044. https://doi.org/10.1016/j.scs.2026.108044

Yohanna, E., & Lim, H. S. (2026). Evaluating regional and seasonal variations of PM2.5 and O3 across Malaysia using XGBoost and SHAP-based interpretability. Chemosphere, 411, Article 145074. https://doi.org/10.1016/j.chemosphere.2026.145074