

## Deep Learning for PM<sub>2.5</sub> Prediction under Zero-Inflated Tropical Rainfall: RNN vs GRU

Ikhsan Yuliyono <sup>1,\*</sup>, Aji Supriyanto <sup>1</sup>

<sup>1</sup> Universitas Stikubank, Indonesia

\* Corresponding author: Ikhsan Yuliyono, Universitas Stikubank, Indonesia  
✉ [ikhsanyuliyono0030@mhs.unisbank.ac.id](mailto:ikhsanyuliyono0030@mhs.unisbank.ac.id)

**Copyright:** © 2026 by the authors

Received: 31 March 2026 | Revised: 12 April 2026 | Accepted: 24 April 2026 | Published: 28 April 2026

### Abstract

Air quality forecasting in maritime tropical regions is challenged by zero-inflated rainfall regimes, where prolonged dry periods are intermittently disrupted by extreme precipitation, generating highly non-linear PM<sub>2.5</sub> dynamics and limiting the effectiveness of conventional predictive models. This study evaluate the predictive performance of Recurrent Neural Networks (RNN) and Gated Recurrent Units (GRU) under such distributional conditions. A quantitative experimental design with a comparative approach is employed using 1,461 daily observations from the Central Java Climatology Station, incorporating rainfall, temperature, and relative humidity as predictors; a chronological data split preserves temporal dependencies, and performance is assessed using MAE, RMSE, MAPE, and R<sup>2</sup> metrics. The results indicate that GRU achieves only a marginal 4.3% reduction in MAE relative to RNN, while both models exhibit substantial predictive failure, as evidenced by negative R<sup>2</sup> values and MAPE exceeding 300%, with predictions collapsing toward the mean and failing to capture extreme pollution events. These findings demonstrate that standard recurrent architectures with conventional loss functions are intrinsically limited in modeling zero-inflated environmental data, contributing empirical evidence on the boundary conditions of deep learning in tropical air quality forecasting and underscoring the necessity for specialized modeling approaches to support reliable early warning systems.

**Keywords:** gated recurrent unit; pm2.5; recurrent neural network; tropical rainfall; zero-inflated data

**To cite this article:** Yuliyono, I., & Supriyanto, A. (2026). Deep Learning for PM<sub>2.5</sub> Prediction under Zero-Inflated Tropical Rainfall: RNN vs GRU. *Edumatic: Jurnal Pendidikan Informatika*, 10(1), 250–259. <https://doi.org/10.29408/edumatic.v10i1.34418>

### INTRODUCTION

PM<sub>2.5</sub> remains a primary indicator of ambient air pollution, responsible for roughly 4.2 million premature deaths annually from cardiovascular and respiratory diseases (Wang et al., 2025). In tropical maritime climates like Central Java, however, air quality forecasting faces a peculiar challenge that temperate models never encounter. Rainfall here is not evenly distributed but "zero inflated": long dry spells are suddenly broken by extreme downpours exceeding 70 mm/day. This intermittency creates volatile PM<sub>2.5</sub> fluctuations through episodic wet scavenging, often with a delayed response of 24 - 48 hours post-precipitation (Ma et al., 2022). Without accurate forecasting, vulnerable populations remain unprotected during pollution surges, and health interventions cannot be deployed proactively. The urgency is clear: context-specific predictive models are needed, but most existing work has been developed for and tested on temperate climates (Slater et al., 2023).



Recurrent Neural Networks (RNN) and Gated Recurrent Units (GRU) are theoretically optimized for capturing non-linear temporal dependencies, with GRU specifically utilizing gating mechanisms to mitigate vanishing gradient issues in volatile sequences. While these architectures consistently outperform traditional statistical models in temperate climates frequently achieving  $R^2$  values above 0.70 their success is predominantly documented in regions with frequent, low-intensity precipitation (Billah et al., 2026). This reliance on stable meteorological signals remains an untested assumption in zero-inflated tropical regimes, where extreme rainfall intermittency presents a fundamentally different predictive challenge that standard sequence-learning frameworks have yet to navigate (Zhang et al., 2023).

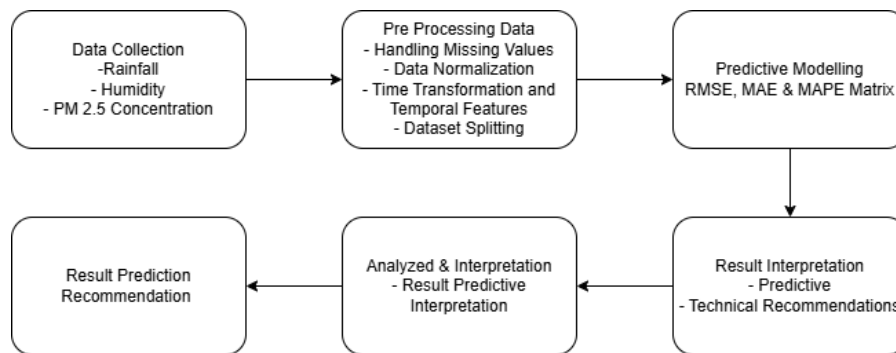
The current state of environmental modeling is characterized by a notable absence of evaluations regarding RNN or GRU performance under zero-inflated tropical rainfall regimes, specifically where prolonged dry periods are punctuated by extreme precipitation events exceeding 70 mm/day. This scarcity of evidence is compounded by a systemic lack of empirical documentation concerning model failure, as negative results are frequently omitted from the literature (Zhang et al., 2025). Extensive review indicates that the response of recurrent neural architectures to the sudden "wash out" effects following extreme downpours in maritime Southeast Asia remains unexamined. Consequently, it remains scientifically undetermined whether generic RNN or GRU frameworks can effectively internalize the non-linear triggers of episodic pollution spikes under such volatile conditions. This oversight transcends geographical specificities, constituting a fundamental theoretical blind spot in contemporary predictive modeling (Shekar et al., 2024).

The present work investigates a twofold scientific oversight regarding the application of deep learning to tropical air quality (Sharma et al., 2023). First, it remains fundamentally unverified whether standard loss functions such as Mean Squared Error (MSE) which inherently promote mean-reverting predictions are technically compatible with the zero-inflated target variables characteristic of maritime tropical climates. Second, there is limited empirical documentation regarding whether the GRU's gating mechanism confers any measurable advantage over the simpler RNN when the primary predictive challenge is distributional skewness rather than long-term dependency (Niu et al., 2023). While prevailing research often assumes that increased architectural sophistication automatically yields superior performance, this study challenges that premise by evaluating the boundary conditions of these models. The scientific justification for this inquiry is that understanding the mechanisms of model failure is as critical to the field's advancement as reporting success (Rajesh & Goswami, 2023). In the absence of documented negative evidence, researchers risk perpetually applying generic architectures to tropical environmental problems, thereby exhausting resources on models that are systemically incapable of capturing extreme pollution peaks or sudden atmospheric wash-out events (Huang & Qian, 2023).

This study demonstrates that standard recurrent neural networks are fundamentally unsuitable for forecasting PM<sub>2.5</sub> concentrations within the high-volatility, zero inflated rainfall regimes of maritime tropical climates. By utilizing 1,461 daily observations to quantify the nonlinear relationship between rainfall and particulate matter, comparing the predictive performance of RNN and GRU architectures, and benchmarking these results against the WHO 24 hours guideline, the research reveals a systemic collapse of predictive utility. The findings indicate that architectural sophistication is secondary to the challenges posed by distributional skewness and the mean reverting bias inherent in quadratic loss functions. Consequently, the primary contribution of this work is the empirical delineation of boundary conditions for deep learning in environmental forecasting, shifting the scientific focus from optimizing generic networks toward the necessity of specialized frameworks such as hurdle models and asymmetric loss functions that can effectively capture episodic wash-out events and critical pollution spikes for reliable early warning systems.

## METHOD

This research adopts a quantitative experimental framework to evaluate the comparative efficacy of Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU) architectures in forecasting PM<sub>2.5</sub> concentrations. The study is structured into a systematic five-stage pipeline: data acquisition, technical preprocessing, architectural implementation, training and validation, and a comprehensive performance critique (Zaini et al., 2022). The primary objective of this design is to assess how standard deep learning structures negotiate the inherent volatility and non-linearities of tropical atmospheric data.



**Figures 1.** Research flow diagram

The primary dataset consists of 1,461 daily observations spanning from January 2022 to December 2024, sourced from the Central Java Climatology Station (BMKG). The target variable is the daily average PM<sub>2.5</sub> concentration ( $\mu\text{g}/\text{m}^3$ ), while the feature space includes critical meteorological predictors: rainfall ( $\text{mm}$ ), air temperature ( $^{\circ}\text{C}$ ), and relative humidity (%). Detailed variable definitions and a representative sample of the raw observations are presented in Tables 1 and 2, respectively (Qing, 2023).

**Table 1.** Description of variables in the dataset

Variable Name	Description	Unit	Data Type
Date	Recording date of daily environmental measurements	dd/mm/yyyy	Date
conct_avg	Daily average concentration of PM2.5 particulate matter	$\mu\text{g}/\text{m}^3$	Float
at_avg	Daily average air temperature	$^{\circ}\text{C}$ (Celsius)	Float
rh_avg	Daily average relative humidity	%(Percentage)	Float
Year	Year of observation	-	Integer
Month	Month of observation (numeric)	-	Integer
Month Name	Month of observation (name)	-	String
Day	Day of the month	-	Integer

**Table 2.** Sample of research dataset (first five days of January 2022)

No	Date	conct_avg	at_avg	rh_avg	Year	Month
1	01/01/2022	14.904.17	27.241.67	78.500.00	2022	January
2	02/01/2022	18.775.00	28.016.67	77.833.33	2022	January
3	03/01/2022	25.537.50	27.800.00	76.125.00	2022	January
4	04/01/2022	26.950.00	28.575.00	70.666.67	2022	January
5	05/01/2022	28.729.17	28.262.50	73.625.00	2022	January

This research employs a quantitative experimental framework to evaluate the comparative efficacy of RNN and GRU architectures in forecasting PM<sub>2.5</sub> levels, utilizing 1,461 daily observations (2022–2024) from the Central Java Climatology Station. The feature space incorporates target pollutant concentrations alongside key meteorological predictors rainfall, temperature, and humidity which undergo rigorous preprocessing to ensure numerical stability and model convergence. Specifically, a  $\log(x + 1)$  transformation is applied to mitigate the severe skewness of zero-inflated tropical rainfall, while a Z-score threshold of  $|Z| > 3$  is implemented to filter technical anomalies without sacrificing the natural volatility of extreme atmospheric events. Furthermore, Min-Max normalization is prioritized over standardization to preserve the structural integrity of pollutant peaks within a  $[0,1]$  range, ensuring that the most critical air quality fluctuations remain represented during the gradient-based optimization process.

## RESULTS AND DISCUSSION

### Results

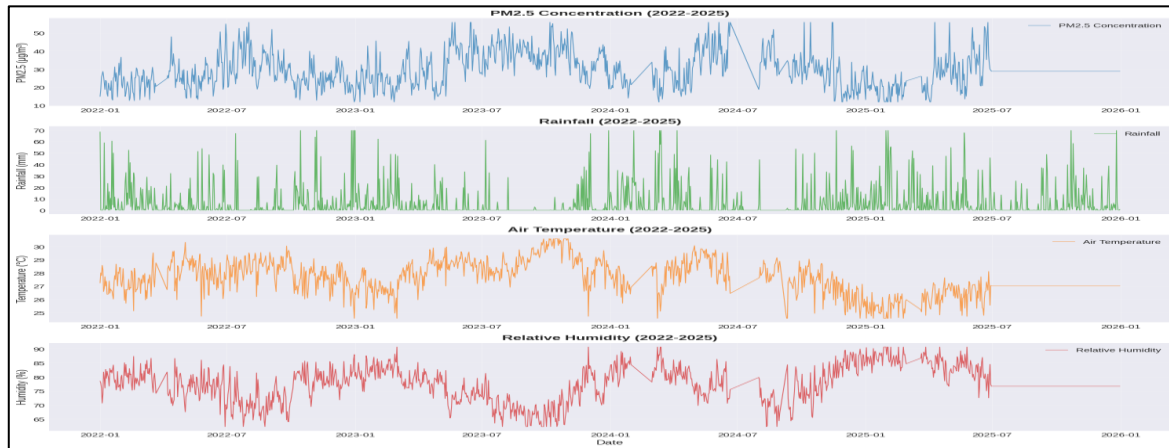
The meteorological profile in Central Java (2022–2025) reflects a dynamic tropical monsoon climate, characterized by a 'zero-inflated' rainfall distribution where a median of 0.00 mm and high standard deviation (13.69 mm) confirm that dry days are frequently punctuated by extreme episodic events. This stochastic pattern directly influences atmospheric cleansing mechanisms, as pollutants are heavily dependent on irregular wash-out events rather than a steady-state environment. While the thermal background remains stable with a mean temperature of 27.57°C, the interplay between persistent high humidity (averaging 77.66%) and unpredictable precipitation cycles creates significant non-linear challenges for predictive modeling. The results in Table 3 show the comprehensive descriptive statistics for these meteorological variables.

**Table 3.** Descriptive statistics

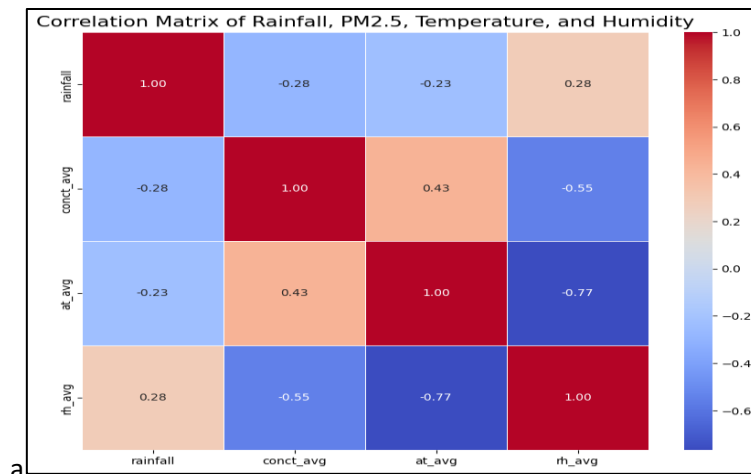
Statistic	Rainfall	Cloud Cover (conct avg)	Temperature (at avg)	Humidity (rh avg)
count	1461.00	1461.00	1461.00	1461.00
mean	6.52	29.46	27.57	77.66
std	13.69	9.28	1.23	5.98
min	0.00	11.78	24.56	62.29
25%	0.00	23.21	26.95	74.25
50%	0.00	28.85	27.43	77.30
75%	5.00	34.50	28.48	81.83
max	69.80	55.88	30.60	90.68

The meteorological profile during the observation period was characterized by a fluctuating cloud cover (mean: 29.46; SD: 9.28; range: 11.78 - 55.88) and relatively stable air temperatures (mean: 27.57°C; SD: 1.23; range: 24.56 - 30.60°C). Relative humidity remained high, averaging 77.66% (SD: 5.98) with a range of 62.29% to 90.68%, where 50% of the data exceeded 77.30%. Integrated temporal analysis reveals that particulate matter concentrations exhibit a distinct inverse response to precipitation. The inverse relationship between precipitation intensity and PM<sub>2.5</sub> concentrations is clearly demonstrated in Figure 2, where high-magnitude rainfall events consistently trigger rapid depletions in pollutant levels. These results indicate that the atmosphere undergoes a significant cleansing process during heavy rain, providing strong empirical evidence for the 'wet scavenging' or 'wash-out' mechanism. Such a correlation underscores that air quality dynamics in Central Java are governed not only by local emission sources but also by the stochastic nature of monsoon precipitation.

Consequently, these findings highlight the inherent complexity of air quality forecasting in maritime tropical climates, as meteorological variability plays a dominant role in determining pollutant dispersion.



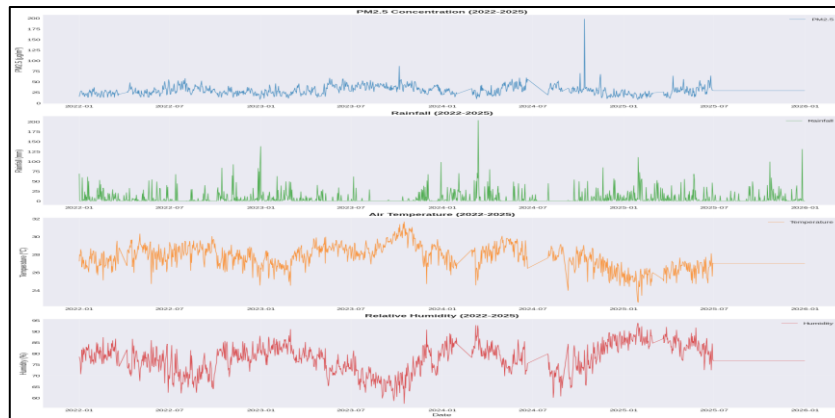
**Figure 2.** Time series plot of PM<sub>2.5</sub> rainfall, temperature, and humidity (2022–2025)



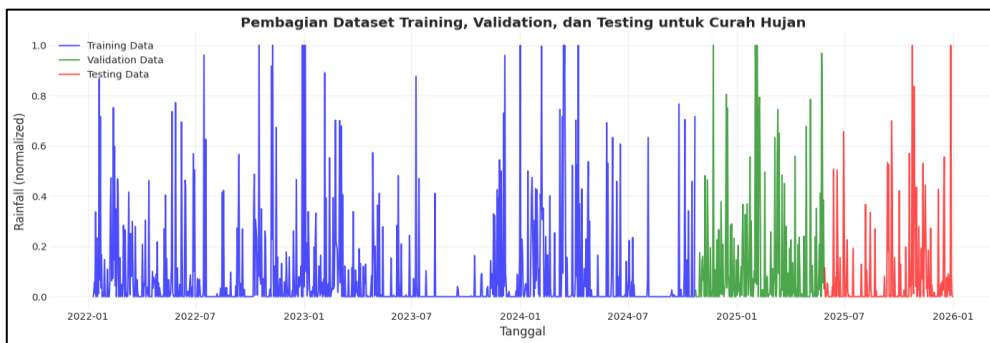
**Figure 3.** Matrix correlation analysis

The correlation matrix presented in Figure 3 reveals that rainfall exhibits only weak associations with the other meteorological variables ( $r$  ranging from  $-0.28$  to  $0.28$ ), while air temperature and relative humidity display a strong inverse relationship ( $r = -0.77$ ). Cloud cover demonstrates moderate positive correlation with temperature ( $r = 0.43$ ) and moderate negative correlation with humidity ( $r = -0.55$ ), collectively indicating that precipitation dynamics in this tropical regime operate largely independently of the thermal and moisture gradients that govern the other atmospheric parameters.

Standardizing the data is a critical step to ensure that larger numerical scales do not disproportionately influence weight updates in RNN and GRU architectures. This linear scaling preserves the fundamental characteristics of the dataset, such as high-intensity tropical rainfall peaks, without introducing structural distortion. As shown in Figure 4, the zero-inflated nature of the precipitation data and the volatile dynamics of meteorological predictors remain invariant post-transformation. This technical optimization improves gradient descent convergence while maintaining the integrity of the underlying non-linear relationships. Finally, the longitudinal dataset was systematically divided into training, validation, and testing subsets to facilitate a rigorous evaluation of model performance, with the chronological allocation of these segments presented in Figure 5.



**Figures 4.** Data normalization



**Figure 5.** Training, validation, and testing data split

The dataset of 1,461 daily observations was reduced to 1,241 sequences using a 7 day sliding window, then split chronologically into 70% training, 15% validation, and 15% testing to prevent data leakage. In comparative performance, GRU achieved marginally lower MAE ( $7.2549 \mu\text{g}/\text{m}^3$ ) than RNN ( $7.5841 \mu\text{g}/\text{m}^3$ ), representing only a 4.3% improvement negligible relative to the PM<sub>2.5</sub> standard deviation ( $13.69 \mu\text{g}/\text{m}^3$ ). Both models exhibited RMSE values ( $\approx 12.3 \mu\text{g}/\text{m}^3$ ) substantially higher than MAE ( $\approx 7.3 \mu\text{g}/\text{m}^3$ ), indicating high error variance and outlier sensitivity. Critically, exceptionally high MAPE ( $>300\%$ ) and negative  $R^2$  values confirm that both architectures fail to explain zero inflated data variance, underpredicting dry-season peaks and missing sudden wash-out events a serious limitation for early warning systems.

**Table 4.** Dataset split statistics

Dataset	Sample Count	Percent (%)	Time
Training	869	70%	2022-2024
Validation	186	15%	2024-2025
Testing	186	15%	2025

The dataset was chronologically split into training (70%), validation (15%), and testing (15%) sets to preserve temporal causality and avoid data leakage. Despite this, GRU achieved only a marginal 4.3% improvement in MAE over RNN, while both models exhibited very high MAPE ( $>300\%$ ) and negative  $R^2$  values. These results indicate a strong mean-reverting bias, with predictions collapsing toward the dataset mean ( $6.52 \mu\text{g}/\text{m}^3$ ) and failing to capture extreme pollution spikes and wash-out events in zero-inflated tropical rainfall conditions.

**Table 5.** Performance comparison of rnn and gru models

Metrik	Model RNN	Model GRU	Difference	Percent (%)
MAE ( $\mu\text{g}/\text{m}^3$ )	75.841	72.549	0.3292	75.841
RMSE ( $\mu\text{g}/\text{m}^3$ )	122.894	123.392	-0.0498	122.894
MAPE (%)	330.50	323.21	7.29	330.50
$R^2$	-0.0042	-0.0124	0.0082	-0.0042

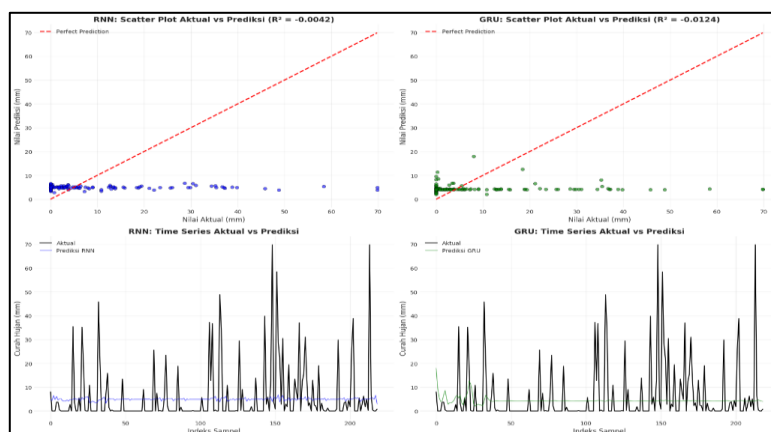
**Figures 6.** Comparison of predicted vs actual value

Figure 6 reveals a significant discrepancy between actual and predicted  $\text{PM}_{2.5}$  concentrations, where both RNN and GRU produced flat prediction lines clustering near the mean value ( $6.52 \mu\text{g}/\text{m}^3$ ), demonstrating a "mean-reverting" bias. This failure is confirmed by negative  $R^2$  values ( $-0.0042$  for RNN,  $-0.0124$  for GRU), persistent underprediction of high-intensity peaks, and overprediction during zero-concentration periods. Residual analysis indicated non-normal error distributions ( $p < 0.001$ ) and heteroscedasticity. Stratified evaluation showed adequate performance at low ( $<20 \mu\text{g}/\text{m}^3$ ) and medium ( $20\text{--}40 \mu\text{g}/\text{m}^3$ ) levels (GRU MAE =  $4.12$  and  $6.89 \mu\text{g}/\text{m}^3$ ), but sharp deterioration for high-pollution events ( $>40 \mu\text{g}/\text{m}^3$ ) with MAE of  $18.47 \mu\text{g}/\text{m}^3$  and MAPE exceeding 340%. The 7 days lag window was optimal, while shorter (3-day) or longer (14-day) windows reduced accuracy.

## Discussion

The empirical results demonstrate a systemic predictive failure of both RNN and GRU architectures under zero-inflated tropical rainfall regimes, evidenced by negative coefficients of determination and Mean Absolute Percentage Errors exceeding 300%. The GRU's architectural sophistication conferred only a marginal 4.3% reduction in error magnitude, a difference that remains operationally inconsequential. Both models defaulted to a stationary prediction near the dataset mean of  $6.52 \mu\text{g}/\text{m}^3$ , failing entirely to capture the episodic spikes and abrupt wash-out events that define air quality variability in Central Java (Areru et al., 2026). In direct response to the research objective, these findings establish that generic recurrent networks are fundamentally incompatible with the stochastic intermittency of maritime tropical atmospheres (Kasipiyawong et al., 2024).

The observed predictive collapse is theoretically attributable to a fundamental mismatch between the Mean Squared Error (MSE) loss function and the zero-inflated distribution of the target variable. Because MSE penalizes residuals quadratically, the optimization process is dominated by the frequent low-magnitude observations, driving the network toward a mean-reverting solution that minimizes global loss at the expense of extreme event fidelity (Bevacqua et al., 2024). This behavior arises from the implicit Gaussian assumption of MSE, which is mathematically invalid for a distribution characterized by a point mass at zero and a

heavy right tail. Consequently, the models treat high-intensity rainfall and pollution spikes as noise rather than as the primary physical signals of wet scavenging and anthropogenic resuspension (Raimondi et al., 2023).

These findings sharply diverge from the preponderance of literature reporting high predictive accuracy for RNN and GRU models in temperate climates, where frequent low-intensity precipitation yields a comparatively stable atmospheric signal (Waqas & Humphries, 2024). The discrepancy is explained by the distinct nature of tropical rainfall regimes, wherein prolonged dry spells are punctuated by stochastic extreme downpours that introduce high-frequency volatility. While temperate models benefit from smooth, continuous meteorological transitions, the zero-inflated intermittency of the maritime monsoon presents a distributional challenge that standard gating mechanisms were never designed to address. This comparison underscores that model success is highly context-dependent and that transferability across climatic zones cannot be assumed without rigorous local validation (Choi et al., 2024).

A particularly unexpected outcome is the negligible performance differential between the RNN and GRU architectures, despite the GRU's theoretical advantages in managing vanishing gradients over extended temporal dependencies (Krishna & Prema, 2023). The empirical data indicate that the primary constraint is not the network's memory capacity but rather the inherent distributional skewness of the target variable. The GRU's reset and update gates, while effective for modulating information flow through time, are powerless to correct an objective function that mathematically incentivizes convergence toward the unconditional mean. This finding challenges the common heuristic that increased architectural complexity necessarily yields improved predictive utility in environmental forecasting applications (Ma et al., 2025).

The primary scientific contribution of this work is the empirical delineation of boundary conditions for recurrent neural networks under zero-inflated tropical atmospheric regimes. By documenting the systemic failure of standard architectures, this study provides essential negative evidence that redirects scholarly attention from architectural refinement toward fundamental considerations of loss function design and distributional alignment (Wang et al., 2025). The findings substantiate the theoretical argument that in environments governed by episodic wet scavenging, the predictive bottleneck is skewness rather than temporal dependency. This work therefore establishes a critical benchmark cautioning against the uncritical deployment of off-the-shelf deep learning solutions in tropical climatological contexts (Pan et al., 2024).

Considering policy and operational implications, the magnitude of prediction error relative to the WHO 24-hour guideline of  $15 \mu\text{g}/\text{m}^3$  renders these models mathematically unsuitable for deployment in Early Warning Systems. The practical consequence is an elevated risk of false negatives, wherein hazardous pollution events are masked by predictions near the annual mean, thereby delaying essential public health interventions (Ma et al., 2025). Furthermore, the reliance on a fixed lag window and a single monitoring station blinds the system to localized anthropogenic resuspension occurring within 24 to 48 hours following rainfall cessation. Environmental agencies operating in tropical regions must therefore abandon expectations that generic pre-packaged models can substitute for context-specific forecasting frameworks that integrate anthropogenic covariates and asymmetric loss functions.

These findings should be interpreted within the context of the study design, which relies on a single monitoring station and does not incorporate anthropogenic covariates into the feature space. Future research should prioritize the construction of multi-station, multivariate datasets that fuse ground-based observations with satellite derived aerosol optical depth and emission inventories (Galvin, 2022). Methodologically, the field requires a transition from symmetric loss minimization toward modeling frameworks that explicitly account for zero-inflated distributions, including the integration of hurdle-based architectures or the application of Tweedie loss functions. This reorientation is crucial to align the theoretical strengths of deep

learning approaches with the practical requirements of robust and reliable air quality forecasting, particularly within the complex dynamics of tropical atmospheric systems.

## CONCLUSION

This study concludes that standard recurrent neural networks are not suitable for forecasting PM<sub>2.5</sub> concentrations under zero-inflated rainfall regimes in maritime tropical climates. Both RNN and GRU exhibit systematic predictive collapse due to mean-reverting bias, failing to capture extreme pollution events. These findings highlight those distributional characteristics, rather than architectural complexity, are the primary limitation, and establish boundary conditions for deep learning in tropical air quality forecasting. Future research should focus on distribution-aware models, such as asymmetric or Tweedie-based loss functions, integration of anthropogenic variables, and multi-source data to improve robustness for early warning systems.

## REFERENCES

- Areru, D. A., Zimale, F. A., & Goshime, D. W. (2026). Integration of machine learning and deep learning models for hydro-meteorological data gap filling and prediction of daily rainfall in the Lake Abaya-Chamo Sub-basin, South Ethiopia. *Journal of Hydrology: Regional Studies*, 65, 103404. <https://doi.org/10.1016/j.ejrh.2026.103404>
- Bevacqua, E., Rakovec, O., Schumacher, D. L., Kumar, R., Thober, S., Samaniego, L., ... & Zscheischler, J. (2024). Direct and lagged climate change effects intensified the 2022 European drought. *Nature Geoscience*, 17(11), 1100-1107. <https://doi.org/10.1038/s41561-024-01559-2>
- Billah, S. M., Haq, H. M., Rahman, A., Biswas, H., & Islam, M. I. (2026). Meteorological drivers of fine particulate matter variability in Dhaka City with a seasonal analysis of temperature, humidity, rainfall, and wind speed. *Discover Environment*, 4(1), 56. <https://doi.org/10.1007/s44274-026-00560-3>
- Choi, H. J., Lee, W. K., & Sohn, S. Y. (2024). PM<sub>2.5</sub> prediction using population-based centrality weight. *Journal of Big Data*, 11(1), 157. <https://doi.org/10.1186/s40537-024-01012-6>
- Galvin, R. (2022). Why German households won't cover their roofs in photovoltaic panels: And whether policy interventions, rebound effects and heat pumps might change their minds. *Renewable Energy Focus*, 42, 236-252. <https://doi.org/10.1016/j.ref.2022.07.002>
- Huang, H., & Qian, C. (2023). Modeling PM<sub>2.5</sub> forecast using a self-weighted ensemble GRU network: Method optimization and evaluation. *Ecological Indicators*, 156, 111138. <https://doi.org/10.1016/j.ecolind.2023.111138>
- Kasipiyawong, J., Gayh, U., & Ghomi, M. R. (2024). The potential of rainwater harvesting and greywater recycling as an alternative domestic water resource in Bahnstadt-Heidelberg, Germany. *Journal of Water, Sanitation and Hygiene for Development*, 14(7), 486-496. <https://doi.org/10.2166/washdev.2024.208>
- Krishna, R., & Prema, K. V. (2023). Constructing and Optimizing RNN Models to Predict Fruit Rot Disease Incidence in Areca Nut Crop Based on Weather Parameters. *IEEE Access*, 11, 110582–110595. <https://doi.org/10.1109/ACCESS.2023.3311477>
- Ma, X., Chen, T., Ge, R., Cui, C., Xu, F., & Lv, Q. (2022). Time series-based PM<sub>2.5</sub> concentration prediction in Jing-Jin-Ji area using machine learning algorithm models. *Heliyon*, 8(9). <https://doi.org/10.1016/j.heliyon.2022.e10691>
- Ma, X., Morawska, L., Zou, B., Gao, J., Deng, J., Wang, X., Wu, H., Xu, X., Wang, Y., Tan, Z., Jiang, N., Shen, Y., Li, D., Gao, J., Fan, Y., & Salmond, J. (2025). Towards compliance with the 2021 WHO air quality guidelines: A comparative analysis of PM<sub>2.5</sub> trends in Australia and China. *Environment International*, 198, 109378.

- <https://doi.org/10.1016/j.envint.2025.109378>
- Niu, M., Zhang, Y., & Ren, Z. (2023). Deep learning-based PM2.5 long time-series prediction by fusing multisource data—A case study of Beijing. *Atmosphere*, *14*(2), 340. <https://doi.org/10.3390/atmos14020340>
- Pan, R., Liu, T., & Ma, L. (2024). A Graph Attention Recurrent Neural Network Model for PM2.5 prediction: a case study in China from 2015 to 2022. *Atmosphere*, *15*(7), 799. <https://doi.org/10.3390/atmos15070799>
- Qing, L. (2023). PM2.5 concentration prediction using GRA-GRU network in air monitoring. *Sustainability*, *15*(3), 1973. <https://doi.org/10.3390/su15031973>
- Raimondi, A., Quinn, R., Abhijith, G. R., Becciu, G., & Ostfeld, A. (2023). Rainwater harvesting and treatment: State of the art and perspectives. *Water*, *15*(8), 1518. <https://doi.org/10.3390/w15081518>
- Rajesh, P. V., & Goswami, B. N. (2023). Climate change and potential demise of the Indian deserts. *Earth's Future*, *11*(8), e2022EF003459. <https://doi.org/10.1029/2022EF003459>
- Riza, O. S., & Nuryadi, A. (2023). Bibliometric study: rainfall Classification-Prediction using machine learning methods. *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, *14*(2), 206-218. <https://doi.org/10.31849/digitalzone.v14i2.16618>
- Sharma, O., Trivedi, D., Pattnaik, S., Hazra, V., & Puhan, N. B. (2023). Improvement in District Scale Heavy Rainfall Prediction Over Complex Terrain of North East India Using Deep Learning. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–8. <https://doi.org/10.1109/TGRS.2023.3322676>
- Shekar, P. R., Mathew, A., Yeswanth, P. V., & Deivalakshmi, S. (2024). A combined deep CNN-RNN network for rainfall-runoff modelling in Bardha Watershed, India. *Artificial Intelligence in Geosciences*, *5*, 100073. <https://doi.org/10.1016/j.aiig.2024.100073>
- Slater, L. J., Arnal, L., Boucher, M. A., Chang, A. Y. Y., Moulds, S., Murphy, C., ... & Zappa, M. (2023). Hybrid forecasting: blending climate predictions with AI models. *Hydrology and earth system sciences*, *27*(9), 1865-1889. <https://doi.org/10.5194/hess-27-1865-2023>
- Wang, G., Ji, Z., Tian, X., Hou, Y., Yang, F., & Ren, F. (2025). Prediction of heavy metal and PM2.5 concentrations in atmospheric particulate matter using key magnetic parameters. *Air Quality, Atmosphere & Health*, *18*(3), 815–825. <https://doi.org/10.1007/s11869-024-01680-6>
- Wang, P., Naso, J., Myers, R., Durney, C. H., Bartolomeu, C., McGuire, A., ... & Lockwood, W. W. (2025). P3. 01.23 The Impact of PM2.5 Exposure on the Genome of Patients With Lung Adenocarcinoma Who Have Never Smoked. *Journal of Thoracic Oncology*, *20*(10), S411-S412. <https://doi.org/10.1016/j.jtho.2025.09.769>
- Waqas, M., & Humphries, U. W. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*, *13*, 102946. <https://doi.org/10.1016/j.mex.2024.102946>
- Zaini, N. A., Ean, L. W., Ahmed, A. N., Abdul Malek, M., & Chow, M. F. (2022). PM2.5 forecasting for an urban area based on deep learning and decomposition method. *Scientific reports*, *12*(1), 17565. <https://doi.org/10.1038/s41598-022-21769-1>
- Zhang, W., Ali, M. Z., Cui, W., Sun, C., Zheng, Z., & Chen, K. (2025). Severe rainfall-induced landslides in Pingyuan County, Guangdong, China, in June 2024: W. Zhang et al. *Landslides*, *22*(8), 2623-2639. <https://doi.org/10.1007/s10346-025-02546-3>
- Zhang, X., Yin, Q., Liu, F., Li, H., & Qi, Y. (2023). Comparative study of rainfall prediction based on different decomposition methods of VMD. *Scientific Reports*, *13*(1), 20127. <https://doi.org/10.1038/s41598-023-47416-x>