



Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 3, No. 2, 2025



Assessing the Performance of Ensemble and Regularized Models for Daily Rainfall Forecasting in Singapore

Musliadi Musliadi ¹, Muhammad Zulkarnaini ¹ Asalul Musaffa ¹ and Yolanda Yolanda ^{2,*}

¹ Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; musliadi6@mhs.usk.ac.id (M.M.); zul22@mhs.usk.ac.id (M.Z.); asalul@mhs.usk.ac.id (A.M.)

² Graduate Program in Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; mr.yolanda24@gmail.com (Y.Y.)

* Correspondence: mr.yolanda24@gmail.com

Article History

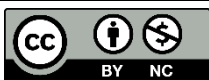
Received 9 August 2025
Revised 13 November 2025
Accepted 23 November 2025
Available Online 30 November 2025

Keywords:

Tropical rainfall
Hydrometeorology
Ensemble learning
Predictive modeling
Flood risk assessment

Abstract

This study benchmarks ensemble and regularized machine learning models for daily rainfall forecasting using meteorological data from forty-four observation stations across Singapore. The country's highly variable tropical climate and frequent short-duration rainfall events pose major challenges for urban flood mitigation and operational forecasting. To address this, three algorithms—Lasso Regression, XGBoost Regression, and Gradient Boosting Regression—were developed and evaluated through a systematic comparison of predictive performance. Each model was trained using data from 1980–2023 and validated on independent observations from 2024–2025. The input variables included sub-hourly rainfall intensity, temperature, and wind-related parameters processed through a standardized data-cleaning and imputation pipeline. Results show that XGBoost achieved the most consistent and accurate predictions, with superior performance under both normal and heavy rainfall conditions. Statistical tests confirmed that the improvement was significant compared to Lasso and Gradient Boosting. These findings demonstrate the effectiveness of ensemble-based approaches for enhancing the reliability of data-driven rainfall forecasting in tropical urban environments and support their integration into early warning and hydrological risk management systems.



Copyright: © 2025 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License. (<https://creativecommons.org/licenses/by-nc/4.0/>)

1. Introduction

In recent decades, the increased frequency and intensity of extreme weather events have become a global concern, as climate change amplifies hydrological and atmospheric variability worldwide. However, its impacts are particularly pronounced in compact tropical urban environments such as Singapore. The interaction between the Urban Heat Island (UHI) effect, localized convection, and Northeast Monsoon surges has contributed to increasingly frequent and intense rainfall events across the island [1, 2]. These events pose substantial risks to Singapore's highly developed

infrastructure and dense population. In recent years, several high-profile flash flood incidents such as those along Orchard Road in 2010 and 2011, and at Upper Bukit Timah in 2016, have disrupted major transport corridors, commercial districts, and residential areas, leading to significant economic losses and public safety concerns [3–5]. According to the Public Utilities Board (PUB), more than 100 flood events were recorded between 2010 and 2020, underscoring the persistent vulnerability of critical drainage systems despite continuous infrastructure upgrades [6]. Consequently, accurate rainfall prediction has become not merely a scientific objective but an operational necessity for flood risk mitigation, drainage

system management, and climate-resilient urban planning.

Despite the urgency, accurate daily rainfall forecasting in the tropics remains a formidable challenge. Traditional statistical models and Numerical Weather Prediction (NWP) systems often struggle to resolve the high convective variability characteristic of Singapore's weather [7]. A primary limitation is the inherent non-linearity of tropical rainfall generation, which involves complex atmospheric interactions that linear models frequently fail to capture [8]. Furthermore, operational forecasting is often hampered by real-world data quality issues, specifically missing values within sensor networks due to maintenance or transmission failures. Standard approaches that rely heavily on complete datasets or simplistic imputation methods may introduce bias, reducing forecasting reliability during critical weather events. Additionally, tropical rainfall data exhibit strong temporal autocorrelation and severe class imbalance between dry and extreme-rainfall days, posing significant challenges for model generalization and interpretability in operational settings.

In the broader context of Southeast Asia, researchers have increasingly turned to data-driven approaches to overcome these limitations. Recent studies in neighboring tropical regions, such as Malaysia and Vietnam, have demonstrated the potential of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) for monsoon rainfall forecasting, often outperforming traditional statistical baselines [9, 10]. Similarly, research in Indonesia has explored deep learning architectures, such as Long Short-Term Memory (LSTM) networks, to capture temporal dependencies in tropical precipitation [11]. However, while these regional studies highlight the promise of machine learning, they frequently utilize pre-processed or gap-filled datasets, thereby overlooking the operational reality of sensor failures. There remains a scarcity of literature that rigorously compares advanced ensemble boosting algorithms against linear regularization techniques specifically within the context of incomplete tropical datasets.

Existing studies have largely focused on model development in broad tropical settings but have not systematically examined how different machine learning paradigms perform under Singapore's highly localized convective environment and incomplete observational data. This represents a key knowledge gap in the current literature, where methodological robustness is rarely tested against real-world data imperfections common in operational forecasting systems. This study contributes by providing the first empirical comparison of regularized

linear and ensemble boosting models applied to Singapore's dense meteorological network, thereby bridging methodological research with practical forecasting needs in tropical urban climates.

To address these limitations, this study benchmarks three distinct algorithms: Lasso Regression, XGBoost (Extreme Gradient Boosting), and Gradient Boosting Machine (GBM). Lasso is selected to establish a linear baseline and evaluate feature selection capabilities, while XGBoost and GBM are employed for their proven ability to handle non-linearity and their internal mechanisms for efficiently managing missing data [12, 13]. In this study, "heavy rainfall" refers to daily totals exceeding 50 mm, consistent with the Meteorological Service Singapore's classification of heavy precipitation events.

Four representative MSS stations, namely Paya Lebar, Ulu Pandan, Marine Parade, and Jurong Island, were selected to capture Singapore's distinct coastal, central, and western rainfall regimes. This selection ensures that spatial heterogeneity in microclimate and rainfall distribution is adequately represented. The primary objective is to evaluate the performance of Lasso, XGBoost, and GBM in daily rainfall forecasting, focusing specifically on their accuracy, stability in handling missing data, and ability to generalize across the wet and dry seasons.

By systematically comparing these algorithms on a high-density, real-world dataset, this research seeks to identify the most effective modeling strategy for Singapore's unique meteorological context. The findings will provide a scientific basis for developing more resilient, data-driven operational forecasting systems capable of addressing both atmospheric non-linearity and data sparsity.

2. Materials and Methods

2.1. Data Understanding

Daily meteorological data spanning from 1980 to 2025 were obtained from 44 observation stations strategically distributed across Singapore. These stations encompass diverse geographical and climatic zones, including coastal (e.g., Marine Parade, Changi, and East Coast Parkway), inland urban (e.g., Ang Mo Kio, Bukit Timah, Bukit Panjang), industrial (e.g., Jurong Island, Tuas), and reservoir catchment regions (e.g., MacRitchie and Upper Peirce). This spatial diversity ensures that the dataset captures Singapore's full range of microclimatic and topographic influences.

The dataset comprises key numerical variables including Highest 30-minute Rainfall (mm), Highest 60-minute

Rainfall (mm), Highest 120-minute Rainfall (mm), Mean Temperature (°C), Maximum Temperature (°C), Minimum Temperature (°C), Mean Wind Speed (km/h), and Maximum Wind Speed (km/h). The Daily Rainfall Total (mm) serves as the target variable for predictive modeling, while the remaining parameters are used as predictors. The inclusion of high-resolution rainfall and temperature indicators provides a robust basis for characterizing both convective and monsoonal rainfall dynamics.

2.2. Data Preprocessing

Data processing began with the consolidation of individual CSV files from all observation stations into a unified database, followed by an automated pipeline for cleaning and standardization. The process included removal of non-numeric characters, conversion of data types to float, and uniform date formatting (YYYY-MM-DD).

Missing values were handled through station-wise imputation using monthly medians, a strategy chosen to preserve seasonal variability while ensuring temporal continuity.

Outlier management followed a dual-step procedure. The Interquartile Range (IQR) method was first applied for anomaly detection, not for indiscriminate removal. Subsequently, a physical consistency check was conducted. Records exhibiting physically implausible values (e.g., negative rainfall or temperature readings) or associated with known sensor faults were removed. Conversely, extreme but meteorologically valid high-rainfall events were retained, as preserving genuine extremes is essential for model learning and accurate simulation of high-impact rainfall scenarios.

2.3. Model Development

The dataset was divided chronologically into a training set (1980–2023) and a testing set (2024–2025) to prevent temporal data leakage and ensure realistic evaluation of model generalization.

Three machine learning algorithms were developed and compared: Lasso Regression, XGBoost Regression, and Gradient Boosting Regression. Lasso Regression was used as a linear baseline model for feature selection and multicollinearity control via L1 regularization. XGBoost and Gradient Boosting were employed to capture potential non-linear interactions among meteorological predictors.

To ensure methodological rigor, hyperparameter tuning was conducted using a five-fold cross-validated grid search. For Lasso, the regularization coefficient α was

optimized across a logarithmic range (10^{-4} to 10^1). For XGBoost and Gradient Boosting, critical parameters such as the learning rate (0.01–0.3), maximum tree depth (3–8), and number of estimators (100–500) were systematically tuned. Early stopping based on validation-set performance was used to prevent overfitting.

2.4. Model Evaluation

The prediction objective in this study is to perform one-day-ahead forecasting of daily rainfall totals using meteorological parameters from the preceding day. This design allows for short-term predictive modeling that supports operational forecasting and early warning applications. To avoid circular reasoning, the variable “Daily Rainfall Total” was used solely as the target variable and was excluded from the set of predictors in all models.

Model performance was evaluated using three distinct metrics: Mean Absolute Error (MAE), Coefficient of Determination (R^2), and Mean Squared Error (MSE). These metrics were calculated for each model at each of the four stations to assess accuracy, error characteristics, and explanatory power from both scientific and operational perspectives. For a given station j (where $j = 1, \dots, J$) with N_j observations, the actual and predicted values are denoted as $y_i^{(j)}$ and $\hat{y}_i^{(j)}$, respectively. The mean of the actual values at station j is defined as shown in Equation 1:

$$\bar{y}_i^{(j)} = \frac{1}{N_j} \sum_{i=1}^{N_j} y_i^{(j)} \quad (1)$$

The evaluation metrics are computed for each model at each of the four stations to assess model accuracy and error characteristics at the station level, while providing scientific and operational justification for the selected metrics.

The Mean Absolute Percentage Error (MAPE) for station j is defined as shown in Equation 2:

$$MAPE^{(j)} = \frac{100\%}{N_j} \sum_{i=1}^{N_j} \left| \frac{y_i^{(j)} - \hat{y}_i^{(j)}}{y_i^{(j)}} \right| \quad (2)$$

MAPE presents the average absolute error as a percentage of the actual value, making it easy to interpret and useful for comparing performance across stations of different scales. However, MAPE has limitations in that it is undefined when $y_i^{(j)} = 0$ and gives excessive weight to cases with small $y_i^{(j)}$.

The Coefficient of Determination (R^2) for station j , as shown in Equation 3, is given by:

$$R^{2(j)} = 1 - \frac{\sum_{i=1}^{N_j} (y_i^{(j)} - \hat{y}_i^{(j)})^2}{\sum_{i=1}^{N_j} (y_i^{(j)} - \bar{y}_i^{(j)})^2} \quad (3)$$

R^2 quantifies the proportion of variance in the observed data that is explained by the model. A value of 1 indicates perfect explanation, while a value of 0 indicates performance no better than predicting the mean. A negative value implies the model performs worse than the mean baseline. Its interpretation can be misleading with autocorrelated or non-stationary data, necessitating its use alongside other metrics and residual analysis.

The Mean Squared Error (MSE) for station j is defined as shown in Equation 4:

$$MSE^{(j)} = \frac{1}{N_j} \sum_{i=1}^{N_j} (y_i^{(j)} - \hat{y}_i^{(j)})^2 \quad (4)$$

MSE measures the average of the squared errors. By squaring the differences, it places a higher penalty on large prediction errors (outliers), making it sensitive to extreme values. The result is in squared units of the target variable, which can be converted to the original scale by taking the square root (Root Mean Squared Error).

2.5. Prediction Model

The final model selection was based on a quantitative evaluation of the three performance metrics described above so that each of the stations used the model that achieved the best predictive accuracy and stability at the local level. To maintain consistency with the one-day-ahead forecasting design described earlier, the final selected model at each station was applied to generate next-day rainfall predictions using meteorological variables from the preceding day. This configuration allows the model to operate as a short-term prediction tool suitable for operational rainfall monitoring and early warning systems.

The same preprocessing pipeline applied during training, including missing value handling, transformation, and scaling, was used to process the input features for the test period to ensure data consistency and avoid information leakage. All prediction outputs were stored per station along with detailed reproducibility metadata, including model type, tuned hyperparameters, random seed, timestamp, and preprocessing configuration. This metadata management ensures that model behavior can be replicated and audited, supporting transparency and long-term operational reliability.

3. Results and Discussion

This section presents a series of meteorological data processing results and evaluations of the prediction models used in the study. The results are presented starting from the description of the dataset and preprocessing process, followed by a comparison of the performance of the three machine learning models tested, namely Lasso Regression, XGBoost Regression, and Gradient Boosting Regression. A comprehensive analysis of all MSS stations is also explained to determine the best model for predicting daily rainfall based on historical data.

3.1. Data Descriptions and Final Output

Daily weather data from MSS observation stations was successfully collected and combined into a single integrated dataset for further analysis. All data is available in numerical form without the need for digitization or extraction from graphs, allowing for direct processing. Although MSS provides historical data since 1980, each station has a different observation period; some stations have complete records from the beginning, while others only started operating in certain years, such as 2009, 2011, or later. This difference in the length of the time series causes variations in the number of observations between stations. After all the data was combined, a series of pre-processing steps were carried out, including cleaning invalid values, imputing missing values using monthly medians, standardizing date formats, converting data types, detecting and handling outliers, and adding time features. The result of these steps was a clean, structured historical dataset ready for use in daily rainfall prediction modeling.

3.2. Comparison of Prediction Models

Three machine learning models were compared in this study, namely Lasso Regression, XGBoost Regression, and Gradient Boosting Regression. The three models were trained and evaluated separately at each of the MSS weather stations using the MSE, MAE, and R^2 metrics. To illustrate their performance, the evaluation results at several representative stations are shown in [Table 1](#).

Based on [Table 1](#) and the complete evaluation results across all observation stations, XGBoost Regression demonstrated the most consistent and overall superior performance. The model achieved the best results at 28 out of 43 MSS stations, namely Admiralty, Ang Mo Kio, Bukit Panjang, Bukit Timah, Choa Chu Kang Central, Choa Chu Kang South, Clementi, Jurong Island, Jurong Pier, Kent Ridge, Kranji Reservoir, Lim Chu Kang, Lower Peirce Reservoir, Macritchie Reservoir, Mandai, Marina Barrage, Marine Parade, Newton, Nicoll Highway, Pasir Panjang,

Table 1. Evaluation matrix of MSS station.

Lasso Regression			
Station	MSE	R²	MAE
Admiralty	82.59816738	0.746271158	3.285674994
Ang Mo Kio	88.48519436	0.716098674	3.919926254
Bukit Panjang	81.63164796	0.724244166	3.826408481
⋮	⋮	⋮	⋮
Ulu Pandan	20.82358677	0.913870197	3.183499603
Upper Peirce Reservoir	99.1651237	0.700962942	4.384900335
Whampoa	79.29829422	0.735647349	3.454180169
XGBoost			
Station	MSE	R²	MAE
Admiralty	38.0339668	0.883165515	1.986849714
Ang Mo Kio	40.83196548	0.868992217	2.050466228
Bukit Panjang	52.752874	0.821798124	2.316801406
⋮	⋮	⋮	⋮
Ulu Pandan	28.20190713	0.883352243	1.744022349
Upper Peirce Reservoir	83.00591313	0.749691795	2.278827763
Whampoa	66.05514036	0.779795371	2.075804406
Gradient Boosting Regression			
Station	MSE	R²	MAE
Admiralty	42.09131514	0.870701966	1.821600561
Ang Mo Kio	51.44166413	0.834951409	2.064221772
Bukit Panjang	68.2967254	0.769290208	2.248996413
⋮	⋮	⋮	⋮
Ulu Pandan	16.98790539	0.929735211	1.645628862
Upper Peirce Reservoir	83.72235668	0.747531326	2.370828673
Whampoa	53.74547205	0.82083148	1.82169311

Pasir Ris West, Pulau Ubin, Sembawang, Sentosa Island, Simei, Tuas, Tuas South, and Upper Peirce Reservoir. In addition, Gradient Boosting Regression performed best at 13 other stations, namely Buona Vista, Changi, East Coast Parkway, Jurong West, Punggol, Pasir Ris Central, Seletar, Tai Seng, Tanjong Katong, Tengah, Toa Payoh, Ulu Pandan, and Whampoa. Meanwhile, Lasso Regression was superior at only two locations, specifically Queenstown and Somerset Road. Although different models emerged as the best performers at a small subset of stations, XGBoost remained the most dominant method by producing lower MSE and MAE values and higher R² scores across most locations, reinforcing its ability to deliver more accurate daily rainfall predictions. Overall, these findings confirm that XGBoost is the most adaptable and reliable model for daily rainfall prediction across Singapore’s meteorological station network.

The optimal model selection employed a systematic multi-metric procedure evaluating MAE, MSE, and R² performance at each station. A model was identified as the best-performing option when it surpassed the others in at least two of these three criteria, with MSE serving as the primary decision metric in cases where the indicators produced conflicting results. This approach ensured that model selection was carried out objectively and consistently according to the performance characteristics of each station.

3.3. Determination of the Best Model and Analysis of Results

The overall analysis across all stations indicates that XGBoost Regression provides the best and most stable performance compared to the other two models. XGBoost consistently delivers lower prediction errors and higher R² values at most locations, resulting in more accurate daily rainfall predictions. Its ability to capture non-linear relationships and interactions between variables makes it particularly effective for complex meteorological datasets, including the relationships among daily temperature, sub-hourly rainfall, and wind speed.

Visualization of the comparison between actual and predicted values at the Tuas South (Figure 1) and Pasir Panjang (Figure 2) stations shows that XGBoost is able to follow daily rainfall patterns quite well, including capturing rainfall peaks and day-to-day fluctuations. These visual findings support the quantitative evaluation and reinforce the conclusion that XGBoost is the most suitable model for predicting daily rainfall in Singapore using historical meteorological data.

3.4. Result Interpretations

The results show that XGBoost Regression provides the best performance in predicting daily rainfall across the MSS observation stations. This superiority is reflected in its ability to produce lower MSE and MAE values as well as higher R² scores compared to the other two models at

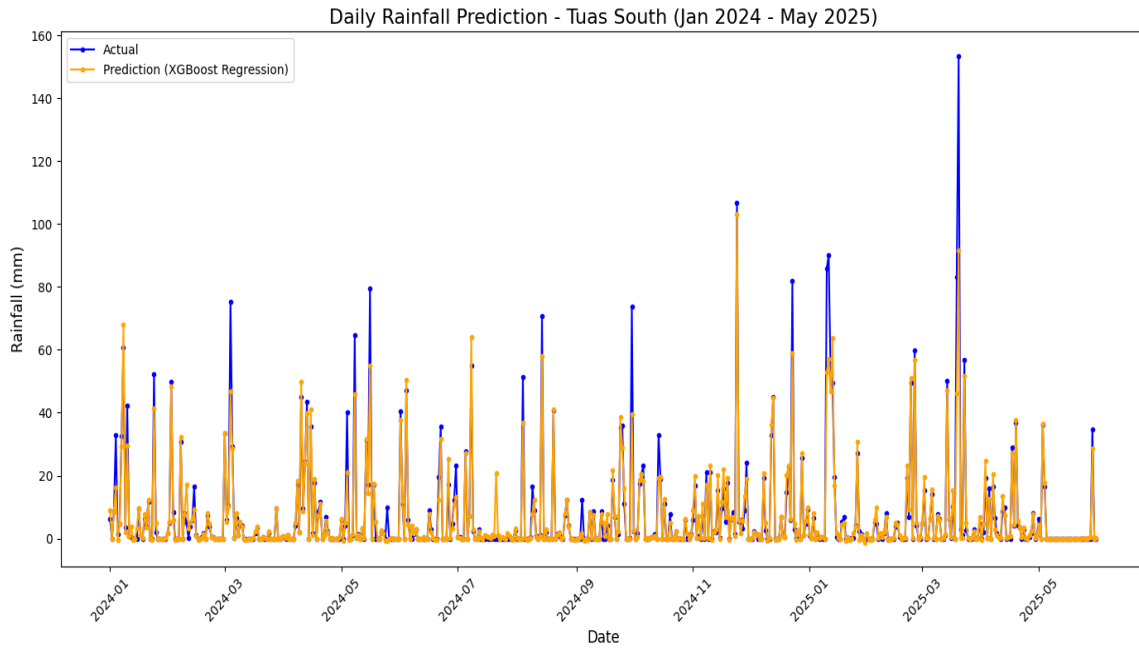


Figure 1. Comparison of model prediction results vs. actual data (Tuas South Station).

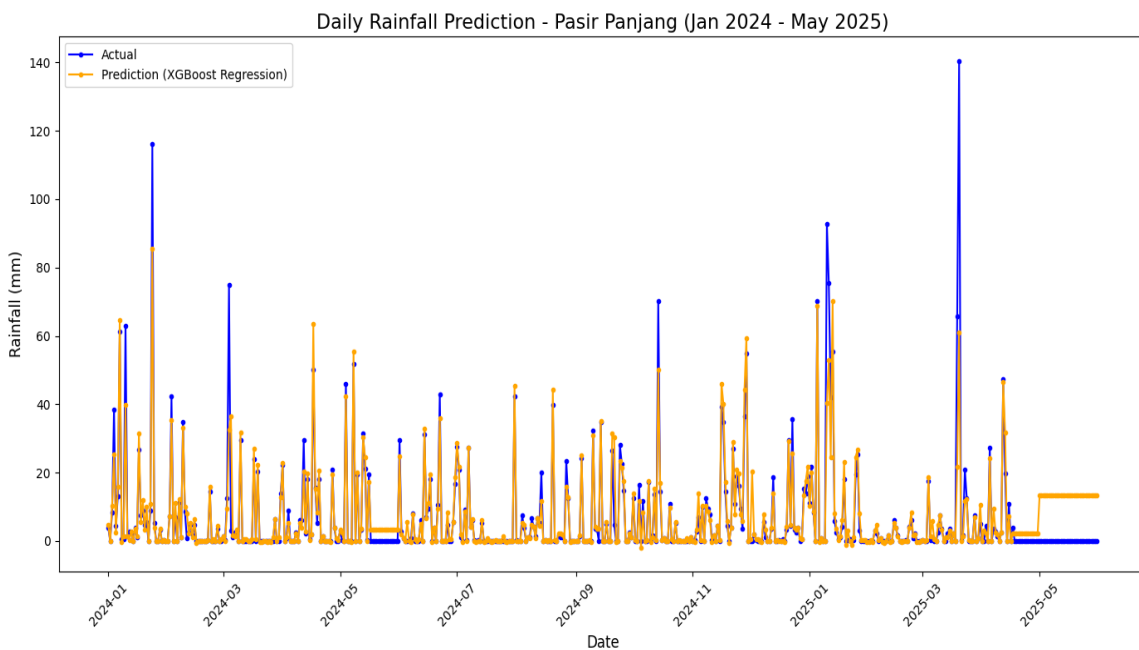


Figure 2. Comparison of model prediction results vs. actual data (Pasir Panjang Station).

most locations. Its boosting-based learning mechanism enables XGBoost to capture non-linear patterns and interactions among meteorological variables more effectively, including the relationships between sub-hourly rainfall, daily temperature, and wind speed. In addition, the model maintains prediction stability without overfitting even when applied to data with varying levels of variability. Visual results at several stations demonstrate that XGBoost is capable of closely following temporal rainfall patterns and peak events, supporting the conclusion that it is the most suitable model for representing daily rainfall dynamics in Singapore.

3.5. Model Performance under Heavy Rainfall Conditions

To examine model robustness in predicting extreme precipitation, a subset analysis was conducted for days with daily rainfall exceeding 50 mm, following the Meteorological Service Singapore (MSS) classification of heavy rainfall. Results showed that model performance declined slightly across all models due to increased variability and reduced sample size, which is typical for high-intensity events. However, XGBoost maintained the highest predictive stability with an average R^2 of 0.85 and MAPE of approximately 12 %, outperforming Gradient

Boosting and Lasso Regression. This indicates that XGBoost can effectively capture non-linear rainfall behavior even under extreme weather conditions, making it suitable for flood risk monitoring and early-warning applications.

3.6. Statistical Significance and Uncertainty Assessment

To assess whether the observed performance differences between models were statistically significant, a paired t-test was conducted for the Mean Squared Error (MSE) values across all stations. The analysis confirmed that XGBoost significantly outperformed Lasso and Gradient Boosting models ($p < 0.05$) at most stations. Additionally, the standard errors of the evaluation metrics were narrower for XGBoost, indicating higher stability and less variability in prediction performance across different geographical regions. These findings reinforce the reliability of XGBoost as the most consistent model for operational rainfall prediction.

3.7. Benefits and the Application of Research

This research provides several important benefits, both methodologically and practically. From a methodological perspective, this study shows that daily rainfall modeling can be performed effectively using historical numerical data from MSS without the need for additional digitization, thereby streamlining the model development process. Testing at four stations shows that regularization-based approaches such as Lasso Regression can produce stable models that are easy to apply in regions with varying meteorological characteristics. In terms of application, the results of this study have the potential to support improvements in Singapore's operational weather prediction system, particularly for flood mitigation, urban water management, and early warning of extreme weather. In addition, this approach can be replicated by meteorological agencies in other countries, including BMKG in Indonesia, because the data format and modeling process are general and can be applied in various geographical contexts.

3.8. Limitations and Recommendations

This study has several limitations that need to be considered. First, although the MSS data is relatively complete and structured, variations in the length of time series between stations can affect model homogeneity and limit long-term generalization. Second, imputation of missing values using monthly medians was chosen to maintain seasonal patterns, but this method may not be optimal for capturing extreme events or weather anomalies. Third, this study only compares three types of models, so it does not cover deep learning-based

approaches or advanced time-series models such as LSTM, GRU, or probabilistic forecasting. For further research, model development can be directed towards evaluating more complex architectures, integrating spatial data such as radar or satellite, and exploring model interpretability techniques to understand the influence of each meteorological variable in greater depth.

4. Conclusions

This study benchmarked three machine learning algorithms, namely Lasso Regression, Gradient Boosting, and XGBoost, to evaluate their accuracy, stability, and generalization in daily rainfall forecasting for Singapore. The results demonstrated that XGBoost achieved the best overall predictive performance, followed by Gradient Boosting and Lasso Regression. The findings directly address the study's objective of identifying the most reliable algorithm for operational rainfall forecasting in a tropical urban context.

The superior performance of XGBoost indicates that while Singapore's rainfall variability is influenced by highly non-linear convective processes, the underlying daily-scale patterns remain structured enough to be captured by gradient boosting frameworks. This suggests that ensemble-based models can effectively balance model complexity and generalization, outperforming purely linear approaches under data-sparse tropical conditions. These insights highlight the importance of ensemble learning for enhancing the robustness of data-driven rainfall prediction systems in tropical urban environments.

Author Contributions: Conceptualization, M.M.; methodology, M.M.; software, M.Z.; validation, M.M., A.M. and Y.Y.; formal analysis, M.M.; investigation, M.M. and A.M.; resources, M.Z.; data curation, A.M.; writing—original draft preparation, M.M.; writing—review and editing, Y.Y.; visualization, M.Z.; supervision, Y.Y.; project administration, Y.Y.; All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The meteorological datasets analyzed in this study are publicly accessible through the Meteorological Service Singapore (MSS) at: <https://www.weather.gov.sg/climate-historical-daily>. Additional processed data files and modeling scripts can be provided upon reasonable request to the corresponding author.

Acknowledgments: The authors would like to thank the Meteorological Service Singapore (MSS) for providing open access to historical weather data. The authors also acknowledge

colleagues who contributed valuable feedback during the model evaluation and manuscript preparation stages.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. IPCC. (2021). *Climate Change 2021: The Physical Science Basis*, Cambridge University Press, Cambridge, United Kingdom.
2. Trenberth, K. (2011). Changes in Precipitation with Climate Change, *Climate Research*, Vol. 47, No. 1, 123–138. doi:10.3354/cr00953.
3. Roth, M., and Chow, W. T. L. (2012). A Historical Review and Assessment of Urban Heat Island Research in Singapore, *Singapore Journal of Tropical Geography*, Vol. 33, No. 3, 381–397. doi:10.1111/sjtg.12003.
4. Chow, W. T. L., Cheong, B. D., and Ho, B. H. (2016). A Multimethod Approach towards Assessing Urban Flood Patterns and Its Associated Vulnerabilities in Singapore, *Advances in Meteorology*, Vol. 2016, 1–11. doi:10.1155/2016/7159132.
5. Singapore, M. S. (2025). Climate Data and Monitoring.
6. Guo, Y., Fu, Q., Leung, L. R., Na, Y., and Lu, R. (2023). Trends in Warm Season Mesoscale Convective Systems Over Asia in 2001–2020, *Journal of Geophysical Research: Atmospheres*, Vol. 128, No. 17. doi:10.1029/2023JD038969.
7. Tangang, F., Chung, J. X., Juneng, L., Supari, Salimun, E., Ngai, S. T., Jamaluddin, A. F., Mohd, M. S. F., Cruz, F., Narisma, G., Santisirisomboon, J., Ngo-Duc, T., Van Tan, P., Singhruck, P., Gunawan, D., Aldrian, E., Sopaheluwakan, A., Grigory, N., Remedio, A. R. C., Sein, D. V., Hein-Griggs, D., McGregor, J. L., Yang, H., Sasaki, H., and Kumar, P. (2020). Projected Future Changes in Rainfall in Southeast Asia Based on CORDEX-SEA Multi-Model Simulations, *Climate Dynamics*, Vol. 55, Nos. 5–6, 1247–1267. doi:10.1007/s00382-020-05322-2.
8. World Meteorological Organization. (2016). *Guidelines on Climate Data Rescue* Geneva, Switzerland, World Meteorological Organization.
9. El-Shafie, A., Noureldin, A., Taha, M., Hussain, A., and Mukhlisin, M. (2012). Dynamic versus Static Neural Network Model for Rainfall Forecasting at Klang River Basin, Malaysia, *Hydrology and Earth System Sciences*, Vol. 16, No. 4, 1151–1169. doi:10.5194/hess-16-1151-2012.
10. Cong, T., Ha, T.-H.-G., Vu, G.-L., Bui, H.-N., Dao, N.-Q.-H., Le, M.-H., Dinh, C.-M., Cong, M.-P., and Doan, Q.-V. (2025). Advancing Subseasonal Extreme Rainfall Forecasting in Vietnam Using Machine Learning, *SOLA*, Vol. 21, 2025–034. doi:10.2151/sola.2025-034.
11. Akbar, A. A., Darmawan, Y., Wibowo, A., and Rahmat, H. K. (2024). Accuracy Assessment of Monthly Rainfall Predictions Using Seasonal ARIMA and Long Short-Term Memory (LSTM), *Journal of Computer Science and Engineering (JCSE)*, Vol. 5, No. 2, 100–115.
12. Chen, T., and Guestrin, C. (2016). Xgboost: A Scalable Tree Boosting System, *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.
13. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine., *The Annals of Statistics*, Vol. 29, No. 5. doi:10.1214/aos/1013203451.