



OPEN

Comparison of machine learning model performance for predicting the climate variables in Johor Bahru, Malaysia

Farid Zamani Che Rose^{1,2✉}, Nur Aqilah Khadijah Rosili³ & Muhammad Fadhil Marsani⁴

Accurately predicting climate variables such as air temperature, humidity and precipitation plays a crucial role in air quality management. This research aims to provide preliminary information that can shed lights to local stakeholders for climate adaptation strategies in Johor Bahru city, Malaysia. Five machine learning models were employed viz. Support Vector Regressions (SVR), Random Forest (RF), Gradient Boosting Machine (GBM), Extreme Gradient Boosting Machine (XGBoost) and Prophet to analyze the 15,888 daily time series climate data in Johor Bahru city, Malaysia. Six climate variables datasets obtained from NASA Prediction of Worldwide Energy Resources (POWER) include Temperature at 2 m (T2M), Dew/Frost Point at 2 m (T2MDEW), Wet Bulb Temperature at 2 m (T2MWET), Specific Humidity at 2 m (QV2M), Relative Humidity at 2 m (RH2M), Precipitation (PREC). Results showed that RF outperforms the other ML models in prediction performance by exhibiting the lowest error for both training and testing data. Superior results are seen for RF in fitting the training data for T2M, T2MDEW and T2MWET with R^2 above 90% demonstrating a strong predictive capability. RF exhibits the lowest error to predict the T2M (RMSE: 0.2182, MAE: 0.1679), T2MDEW (RMSE: 0.2291, MAE: 0.1750), T2MWET (RMSE: 0.1621, MAE: 0.1251), QH2M (RMSE: 0.3502, MAE: 0.2701) and RV2M (RMSE: 1.4444, MAE: 1.1090). RF shows particularly strong Nash–Sutcliffe efficiency (NSE) scores up to 0.94 in the training phase, especially for temperature-related variables indicating high explanatory power and stability. In contrast, SVR demonstrates superior generalization in the testing phase, with the highest Kling-Gupta Efficiency (KGE) value (0.88) confirming its reliability in out-of-sample forecasting. The findings of this research provide transparent, data-driven insights that can inform policymakers and guide the development of robust public policies and strategic investments in Johor Bahru.

Keywords Machine learning, Prediction, Temperature, Humidity, Precipitation, Adaptation, Climate

The climate is changing worldwide and affecting all regions of the world. Temperature fluctuations, rainfall, precipitation, melting glaciers, sea level rise, and frequent extreme weather events are examples of changes that affect millions of people and animals around the world. Recent studies have shown a 1.14-degree Celsius increase in temperature in Peninsular Malaysia between 2000 and 2019¹. This issue poses a significant challenge, particularly for serious risks to both human and natural systems, especially in regions where weather patterns are highly sensitive to monsoonal effects. Accurate prediction of climate variables is therefore crucial for climate adaptation, resource planning, and risk mitigation.

Numerous studies and reviews have been undertaken to explore the application of machine learning (ML) models to predict climate variables such as in^{2–9}. These studies have contributed to examining the nonlinear characteristics, chaotic environmental processes, such as soil, land, and Earth temperature. The application of Extreme Gradient Boosting (XGBoost) model has been widely used in many fields such as climate predictions^{2–4} classification of COVID-19 patients^{5,10} and agriculture sector such as in¹¹. Details of XGBoost algorithm can be found in¹². The algorithm was originally derived from gradient boosting decision tree. This

¹Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia. ²Institut Penyelidikan Matematik (INSPeM), Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia. ³Faculty of Computing and Engineering, Quest International University, 30250 Ipoh, Perak, Malaysia. ⁴School of Mathematical Sciences, Universiti Sains Malaysia, 11800, USM Minden, Pulau Pinang, Malaysia. ✉email: faridzamani@upm.edu.my

ML model can be considered as the most flexible model since it can be used for regression and classification purposes. The Random Forest (RF) model has one common fundamental characteristic with Gradient Boosting Machine (GBM) and XGBoost which is decision making based on decision tree. It started with a root node, having multiple levels of child nodes until then end called leaf node (decision node). Numerous applications of RF model in climate variables prediction^{13–15}. Facebook's Prophet algorithm also has been widely used in prediction climate variables^{16,17}.

Machine learning algorithms generally outperform traditional time series models in forecasting, particularly when combined in hybrid models or enhanced with nonlinear transformations¹⁸. Previous studies have demonstrated the utility of machine learning models in environmental and climate forecasting applications. However, several limitations in the existing literature justify the need for further exploration. First, prior research often evaluates only one or a narrow set of models, limiting understanding of their comparative strengths and generalizability¹⁹. Second, while tree-based and kernel methods have proven effective, their performance in tropical, monsoon-influenced urban regions such as Johor Bahru remains underexplored^{20,21}. Third, Prophet's effectiveness under conditions of high interannual variability or zero inflated time series, such as daily rainfall, has shown to be limited in recent studies¹⁷. While traditional statistical models such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) are commonly used in time series forecasting, their assumptions of linearity and stationarity limit their effectiveness when applied to climate data characterized by nonlinear trends, abrupt changes, and multivariate dependencies. Prior study has shown that machine learning models often outperform classical approaches in such contexts. Recent study by²² reviewed 94 studies on short-term flood forecasting and found that ML approaches, such as RF, Long Short-Term Memory (LSTM), and hybrid models, consistently achieved higher predictive accuracy than traditional statistical models.

Forecasting climate variables is necessary for effective ecosystem management in terms of energy, cost, time, and resources²³. This research is driven by the critical need to address the impact of climate change on in Johor Bahru as urban area in Peninsular Malaysia that is common for flash floods to occur²⁴. Johor Bahru has experienced a notable increase in temperature over recent decades, accompanied by more intense dry and wet spells. These shifts highlight the growing need for localized climate prediction capabilities to support adaptation efforts. Addressing these challenges is crucial, and this study aims to fill this gap by exploring several pertinent research questions:

1. How do various machine learning (ML) models perform in predicting the climate variables in Johor Bahru?
2. Which ML model is most effective for climate time series predictions in Johor Bahru?

To address the questions outlined above, the present study evaluates the performance of five widely adopted machine learning models viz. Support Vector Regression (SVR), RF, GBM, XGBoost, and Prophet for predicting daily climate variables in Johor Bahru, Malaysia. These models were carefully selected to represent a diverse set of methodological approaches, each offering distinct advantages for time series forecasting. SVR, which is based on kernel methods, is particularly effective at capturing complex nonlinear patterns within relatively small datasets. RF, GBM, and XGBoost are all ensemble learning algorithms that build multiple decision trees to enhance prediction accuracy. These models are well-suited for handling noisy data, uncovering intricate relationships among variables, and providing insights into the relative importance of input features. Prophet, developed by Facebook, is a decomposition-based model that separates trend and seasonality components. It is especially useful for producing interpretable forecasts in datasets with strong temporal patterns.

Although these models are individually well-established, their comparative evaluation using a long-term climate dataset in a tropical urban region such as Johor Bahru has not been extensively explored. To our knowledge, no previous study has conducted such a long-term, multi-model evaluation of climate prediction performance specifically tailored to Johor Bahru's tropical climate conditions. This study represents one of the first comprehensive assessments of their performance using more than four decades of daily climate data from NASA's POWER database. In contrast to previous research that primarily emphasizes newly developed algorithms, this study adopts a balanced approach by evaluating both established classical models such as SVR and RF and more recently introduced models such as XGBoost and Prophet. This strategy enables a broader understanding of their strengths and limitations when applied to tropical, monsoon-affected climates. Such benchmarking is vital not only for identifying the most suitable model for regional forecasting but also for informing the design of predictive frameworks that can be adapted to other cities in Southeast Asia. By analysing differences in prediction accuracy, model transparency, and computational efficiency, this work offers valuable guidance for supporting climate risk assessments, regional adaptation strategies, and the development of localized early warning systems. By precisely predicting climate variables, this study supports the efficient allocation of resources, ensuring sustained productivity and reducing the risks associated with flood²⁵.

The rest of this paper is as follows: the area of study, dataset and methods section presents the study area, data explanation and describes the concept and algorithm of the applied ML models, along with a thorough analysis of their statistical performance. Then, the results and discussions section present descriptive statistical analyses and machine learning performance in predicting the climate variables in Johor Bahru. Finally, the conclusion summarizes the preceding discussion and outlines future research objectives.

Area of study, dataset and methods

Malaysia's proximity to the equator gives it a tropical climate, characterized by hot and humid conditions throughout the year. Due to its location close to equatorial line, it means Malaysia does not having a four season but only sunny and rainy days with temperature ranging from 23 to 32 Celsius²⁶. Generally, Malaysia has two regions: West Malaysia (Peninsular or Malaya) and East Malaysia (Borneo). Johor state is in south Peninsular

Malaysia which covers an area of 19,210 square kilometer linked with Singapore by a causeway. The specific location of study is the capital city, Johor Bahru situated at a latitude of 1.4927 and a longitude of 103.7511 as shown in Fig. 1. The selection of the study area due to its increasing vulnerability to climate-related hazards, particularly flooding. Located in a tropical monsoon zone near the equator, the city experiences high rainfall variability, elevated humidity, and complex seasonal transitions influenced by both northeast and southwest monsoons. These climatic characteristics, combined with rapid urbanization, have heightened their susceptibility to extreme weather events. Notably, in early 2025, Johor Bahru faced severe flooding due to continuous rain that displaced over 4,000 residents and disrupted infrastructure. Johor Bahru has well known history of flash floods²⁴ occurred and severe drought in 1990, 1997, 2005, and 2010–2014 leading to water shortage²⁷.

This study used 15,888 daily climate time series data from 1 January 1981 to 30 June 2024. The dataset was obtained from NASA Prediction of Worldwide Energy Resources (POWER) website which are freely accessible at <https://power.larc.nasa.gov/data-access-viewer/>. The website provides daily gridded climate data at a spatial resolution of approximately $0.5^\circ \times 0.5^\circ$ (about $55 \text{ km} \times 55 \text{ km}$ at the equator). This level of resolution is appropriate for city-scale analysis and has been validated in prior studies for use in tropical and subtropical regions. The dataset has undergone quality control and preprocessing to address common data quality issues. As part of the data preparation process, we performed an initial screening of the dataset, including visual inspection and descriptive statistical analysis, to identify any missing or inconsistent values.

The dataset was partitioned into two where 80% was used for training and the rest utilized for testing purposes. This method allows for assessing how well each model predicts unseen data, which is essential in

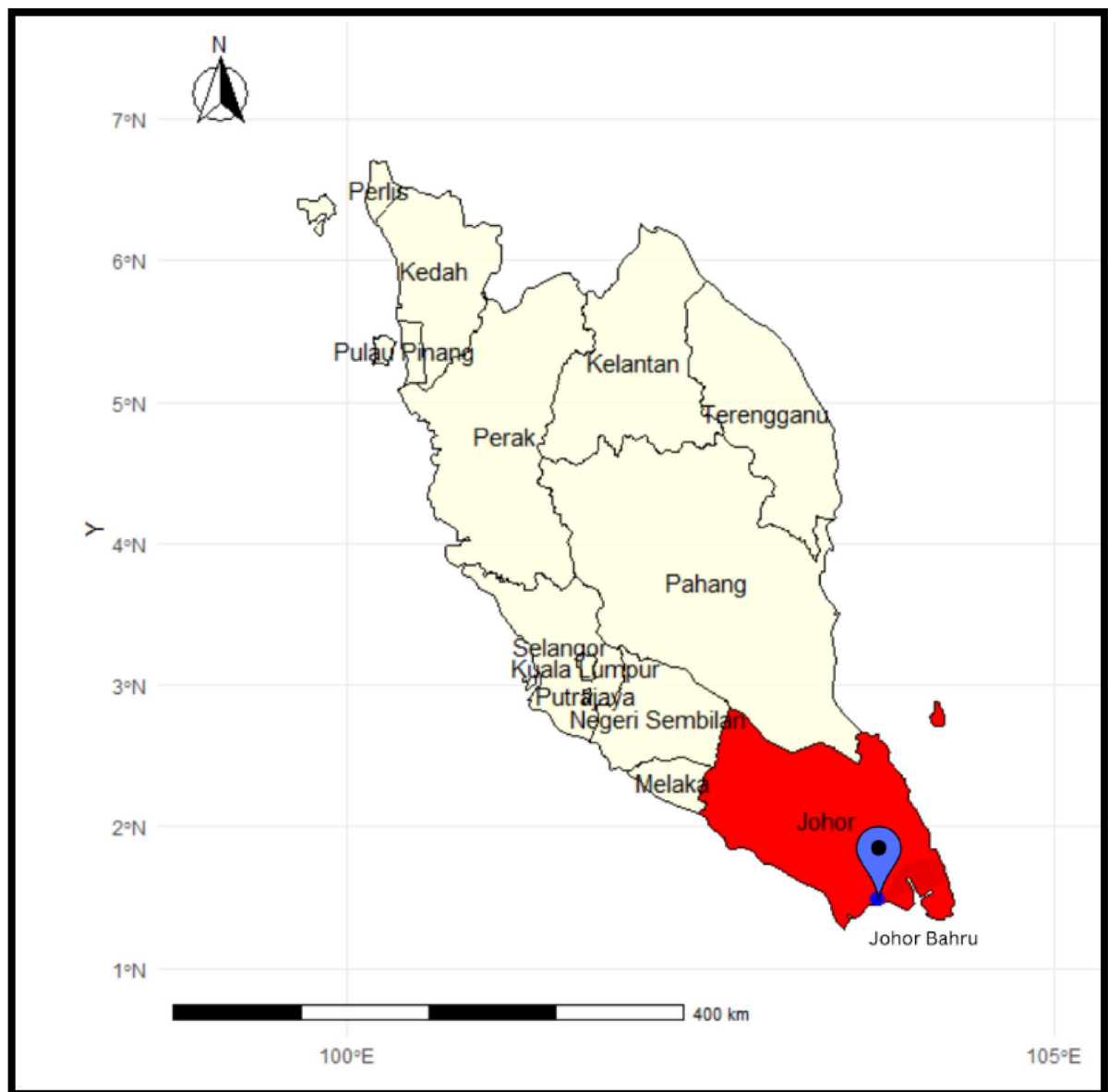


Fig. 1. Location of the study area (Johor Bahru, Malaysia) with a highlighted geographical boundary.

practical forecasting applications. Traditional k-fold cross validation was not used in this study, as it may lead to data leakage and unrealistic performance estimates when applied to time series data. Instead, the temporal sequence of the observations was preserved to maintain the integrity of the time-dependent structure. Hence, we employed rolling-window cross validation or time series cross validation techniques to further enhance model robustness while respecting the temporal nature of the data. This approach ensures that model evaluation reflects real-world forecasting conditions, where future values must be predicted based on past observations. Subsequently, five different supervised ML models were employed to predict daily climate variables in Johor Bahru. Finally, the performance of each model was assessed based on performance metrics. Table 1 presents the summary of selected climate variables used indicating the abbreviation, SI unit, and period of study.

This section provides brief technical details about each model. A graphical description of the procedure is presented in Fig. 2 below.

Prophet

The Facebook team has developed an open-source algorithm to analyse and forecast the time series data. It offers strength in analysing trends, seasonal changes, logarithmic data, robustness to outliers, and handling missing data. Details of Prophet's algorithm can be found in²⁸. Prophet represents the time series data as an additive model consisting of trend, seasonality, and holidays as follows:

$$y(t) = T(t) + S(t) + H(t) + \varepsilon_t \quad (1)$$

where $T(t)$ is the trend function, $S(t)$ is the seasonality function(s), $H(t)$ is the holidays function or temporary event that changes the mean of time series and ε_t represents any changes in time series data cannot be captured by the model which assumes to be independent and normally distributed.

Support vector regressions (SVR)

Support Vector Machine (SVM) which works by identifying the optimal hyperplane for classification purposes. However, for regression purposes the SVR algorithm works by identifying the optimal hyperplane which fits the data closely at an acceptable degree of error. Pioneering work on SVM and SVR has been carried out by²⁹. Generally, the function $f(x)$ is used to identify the optimal association between input data and output data. The function of can be described as follows:

$$f(x) = wx + b \quad (2)$$

where $f(x)$ is the predicted output data, x is the input data, w is the weight vector, and b represents the bias component. In SVR algorithm, the objective is to determine the optimal value of w and b to minimize the error between predicted output data and true data.

Gradient boosting machine (GBM)

Gradient Boosting Machine (GBM) works by building a series of small, simple models (weak learners), typically decision trees, one at a time. Each new model is trained to correct the errors made by the previous models, and the process continues iteratively. The final prediction is made by combining all these models, resulting in a stronger, more accurate overall prediction³⁰.

Extreme gradient boosting (XGBoost)

The development of XGBoost is an extension from GBM. The key enhancement of prediction accuracy in XGBoost algorithm is the internal algorithm that combines the output from multiple individual decision tree in sequence, with each new tree correcting the errors of the preceding tree. Assume that F is a basic tree model and M is a total of multiple trees, then:

$$\hat{y}_i = \sum_{m=1}^M f_m(x_i), f_m \in F \quad (3)$$

Climate variables	Unit	Period
Temperature at 2 m (T2M)	°C	1 Jan 1981–30 June 2024 Train data (80%) Test data (20%)
Dew/Frost Point at 2 m (T2MDEW)	°C	
Wet Bulb Temperature at 2 m (T2MWET)	°C	
Specific Humidity at 2 m (QV2M)	g/kg	
Relative Humidity at 2 m (RH2M)	%	
Precipitation (PREC)	mm	

Table 1. Dataset on climate variables used in the study, including their abbreviations, SI units, study period, and the train-test split (80–20%).

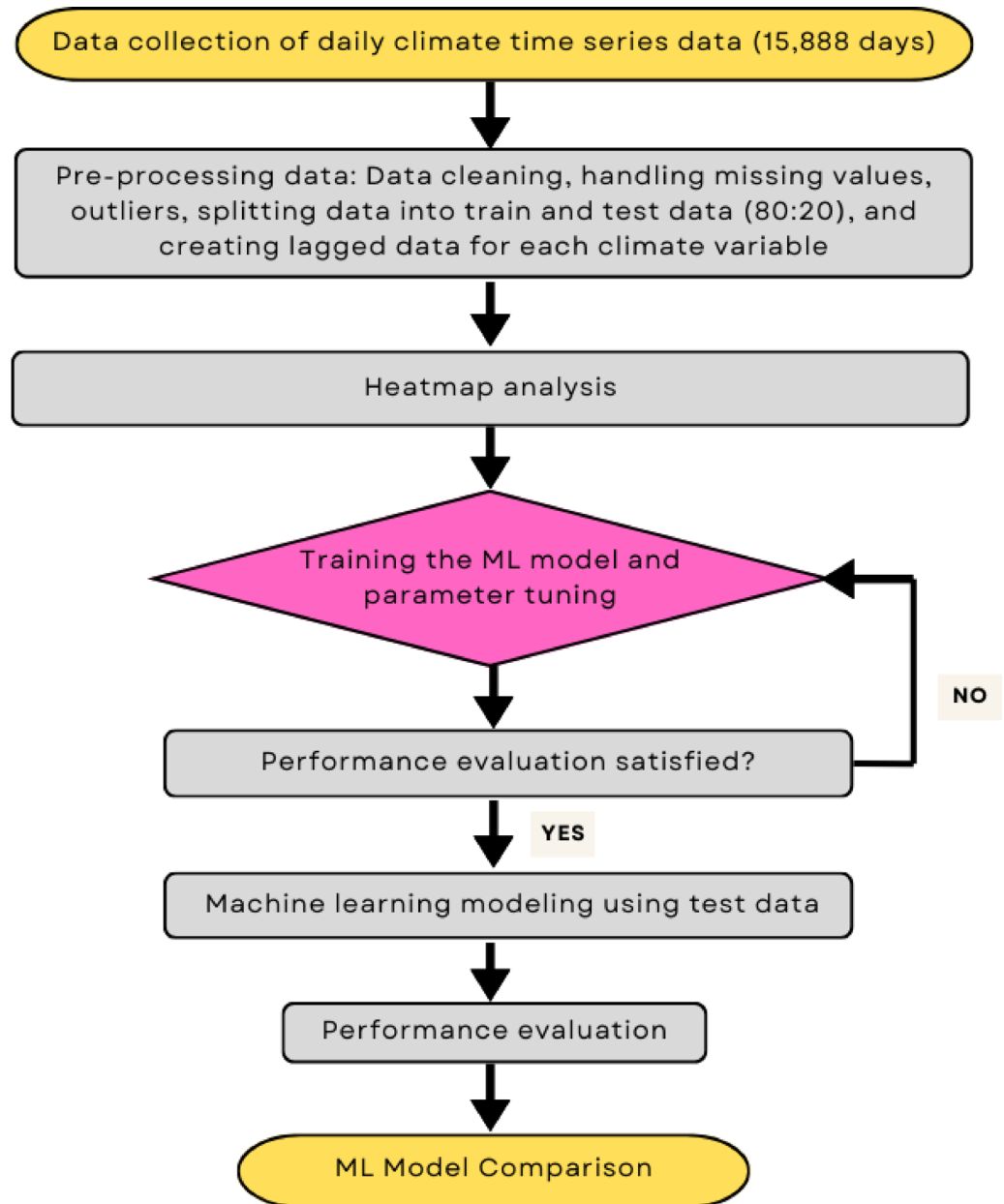


Fig. 2. Flowchart of the methodological framework used in the study. It outlines the stages from data acquisition, preprocessing, and variable selection to model development using machine learning models.

Generally, the objective function of XGBoost model is as follows:

$$obj(\theta) = \sum_i L(\hat{y}_i, y_i) + \sum_i \Omega(f_k) \quad (4)$$

where L represents the loss function or the error function obtained from difference between predicted value and the true value, Ω is the regularized function to avoid overfitting.

Random forest (RF)

RF is a non-parametric model that is available for regression and classification purposes. Previously, the fundamental concept in GBM and XGBoost was based decision trees. Instead of having single trees, the random forest algorithm growing N different trees by randomizing the input. The key ideas in RF algorithms are randomly selecting bootstrap samples from the training dataset, randomly selecting the vector which controls the growth of N trees and randomly selecting the subset of input at each node to determine the split. This process

occurs iteratively until the final decision is made at the final terminal model. Pioneering work on this subject was carried out earlier by³¹.

Key parameters of machine learning model

Referring to Table 2, we trained all the models using the caret package in R with hyperparameter tuning via 5 points grid (tunelength=5). Our approach applied time series specific cross validation strategy using the timeslice method, which respects the temporal order of observations. This study includes preprocessing steps including centering and scaling for SVR, while tree-based models require no normalization. A separate implementation for Prophet since the model requires its default configuration using additive seasonality and automatic changepoint detection.

Performance evaluation

The ML models were evaluated using four indicator metrics which are root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), Nash–Sutcliffe efficiency (NSE), Kling–Gupta Efficiency (KGE), Taylor diagrams and Friedman’s test. The equation of performance metrics are as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

$$NSE = 1 - \frac{MSE}{\sigma^2} \quad (8)$$

$$KGE = 1 - \sqrt{(r - 1)^2(\alpha - 1)^2(\beta - 1)^2} \quad (9)$$

where n is the number of data point computed, y_i is the true value, \hat{y}_i is the predicted value, \bar{y} is the average true value. RMSE is used to illustrate the difference between predicted and true values. Lower values of MSE and RMSE indicate better performance in prediction by ML models. The R^2 coefficient represents the degree of fit between the predicted and true values. R^2 value closer to 1 indicating that the model is a good fit. NSE is closely related to MSE in Eq. (8) divided by σ^2 , the variance of true value. KGE introduced by³² consists of three distinctive components representing correlation, bias, and a measure of relative variability in the simulated and observed values. The r is the linear correlation between observations and simulations, α is a measure of the flow variability error, and β denoted as bias.

This study employed Friedman’s test to compare the performance of ML models using statistical tests. The Friedman test statistic is based on the average ranked (AR) and computed as follows:

$$\chi_F^2 = \frac{12D}{K(K+1)} \left[\sum_{j=1}^K AR_j^2 - \frac{K(K+1)^2}{4} \right] \quad \text{where } AR = \frac{1}{D} \sum_{i=1}^D r_i^j \quad (10)$$

where D denotes the number of data sets used in the study, K is the total number of models. Each model is ranked from 1 = best to k = worst model for each data set. If the value of χ_F^2 is large enough, then the null hypothesis that there is no difference between the techniques can be rejected. Following a significant Friedman test result ($p < 0.05$), we applied the Nemenyi post hoc test to perform pairwise comparisons between models.

Model	Method in R package	Tuned parameters
SVR	svmradial	Sigma, c (via tunelength=5), radial basis kernel
RF	rf	mtry, (number of variables tried at each split); ntree=500 (default in randomForest package)
GBM	gbm	n.trees, interaction.depth, shrinkage, n.minobsinnode
XGBoost	xgbtree	nrounds, maxdepth, eta, gamma, colsample_bytree
Prophet	prophet	No tune in this setup

Table 2. Hyperparameter settings applied to the five machine learning models. The selected parameters were tuned to optimize predictive performance across the different climate variables.

The Nemenyi test identifies which model pairs differ significantly by comparing their average ranks, with a critical difference (CD) calculated based on the number of models and datasets. The test statistic value is based on the CD between average ranks, computed as:

$$CD = q_{\alpha} \sqrt{\frac{k(k+1)}{6N}} \quad (11)$$

where q_{α} is the critical value from the Studentized range distribution (based on k treatments at significance level, α). If the absolute difference between two models' average ranks exceeds the CD, their performances are considered significantly different. This approach is widely used in benchmarking scenarios involving multiple models across multiple datasets.

Results

This section presents the findings from the analysis of climate variables in Johor Bahru. The parameter is important to determine the state of humidity in the air. In Malaysia, the Southwest Monsoon (late May to September) and the Northeast Monsoon (November to March) significantly influence weather patterns, including dew point temperature, wet-bulb temperature, humidity, and precipitation³³. The analysis begins with a visualization of converted average monthly time series data, which is illustrated in Fig. 3. The heatmap analysis reveals distinct seasonal patterns in Malaysia, with T2M tending to peak during the inter-monsoon periods, specifically from March to June and again from September to October³⁴. These peaks correspond with the transitional phases between the monsoon seasons, where temperature increases are more pronounced. During this time, there is often intense convection leading to afternoon thunderstorms and higher humidity levels across the country. Dew point and wet-bulb temperatures can remain high due to the moist atmosphere, and short but intense rainfall is common. The heatmaps clearly reflect these seasonal fluctuations, showing higher temperatures during these periods. This finding aligns with our expectations that southern Peninsular Malaysia has been experiencing more frequent heatwaves since the year 2000, which have increased the heatwave-affected area by 8.98 km² per decade and extended its duration by 1.54 days per decade²⁶. Furthermore, the region has been declared under multiple Heatwave Level 1 alerts by local government agencies due to dry conditions and reduced rainfall. This condition was confirmed by³⁴ which reveals that Johor experienced extreme wet and dry years, leading to drought and flood incidents. Less rainfall over the period leading to severe dry seasons in Malaysia^{35,36}.

To assess the presence of trend and seasonal components in the climate data, we performed the Seasonal-Trend decomposition using Loess (STL) on all six variables. Each panel in Fig. 4 shows the time series decomposition into observed data, seasonal component, trend, and remainder. The decomposition revealed strong and regular annual seasonality in temperature-related variables (T2M, T2MDEW, T2MWET) and humidity variables (RH2M, QV2M), supporting the use of additive seasonal models such as Prophet. These variables also displayed smooth, interpretable long-term trends. In contrast, PREC exhibited irregular spikes and a weakly defined seasonal component, with large residual fluctuations indicating abrupt changes and stochastic variability.

Next, an important note must be made regarding the performance of five ML models using training data viz. RF, SVR, GBM, XGBoost, and Prophet in Johor Bahru. The evaluation focuses on the R^2 values for each model using training and testing presented in Figs. 5 and 6 respectively. Experimental results reveal the RF and SVR offer performance advantages in fitting the training and testing data respectively. Superior results are seen for RF in fitting the training data for T2M, T2MDEW and T2MWET with R^2 above 90% demonstrating a strong predictive capability. Similarly, for QV2M and RH2M, the RF also demonstrates a strong predictive capability with 81% variance in the training data can be explained. Meanwhile, the SVR also offers promising performance in fitting the testing data. However, the R^2 obtained are slightly lower than in training data. This is mainly caused by the limited number of observations as the data were partitioned into 80:20 for training and testing data respectively.

Results illustrated in Tables 3 and 4 show that the performance metrics for five ML models using training and testing data respectively. In the course of this work, we discovered that the RMSE and MAE reported in Tables 3 and 4 for PREC are higher than the other climate variables. Our most intriguing finding is that the RF outperforms the other ML model by exhibiting the lowest error for both training (RMSE: 5.4276, MAE: 2.9374) and testing (RMSE: 9.7572, MAE: 5.6366) data. This is an interesting finding, and it could be hypothesized that the precipitation data in Johor Bahru are imbalanced (contains zero values) indicating dry days and no rainfall. Hence, the ML models struggle to perform in predicting the precipitation data as compared to other climate variables. On the other hand, the results obtained in Tables 3 and 4 appear to tally with our expectations of low RMSE and MAE for other climate variables i.e. T2M, T2MDEW, T2MWET, QH2M and RV2M. Besides, these results demonstrate the potential superiority of RF over established ML models used to predict the T2M (RMSE: 0.2182, MAE: 0.1679), T2MDEW (RMSE: 0.2291, MAE: 0.1750), T2MWET (RMSE: 0.1621, MAE: 0.1251), QH2M (RMSE: 0.3502, MAE: 0.2701) and RV2M (RMSE: 1.4444, MAE: 1.1090).

In addition to R^2 , RMSE, and MAE, the evaluation using KGE, and NSE further clarifies the relative performance of the machine learning models. In the training phase (See Table 3), RF consistently yielded the lowest MAE across all six variables, demonstrating minimal average error. Its strong performance is also reflected in high NSE values, indicating reliable replication of observed patterns, particularly for temperature-related variables where NSE exceeded 0.93. The KGE values above 0.85 for RF and XGBoost further confirm their robust ability to capture bias, variability, and correlation. On the other hand, SVR produced the lowest MAE for all variables except QV2M, suggesting it generalizes well across climate variables in the testing phase (See Table 4). SVR also showed the highest KGE values for five out of six variables, indicating superior predictive

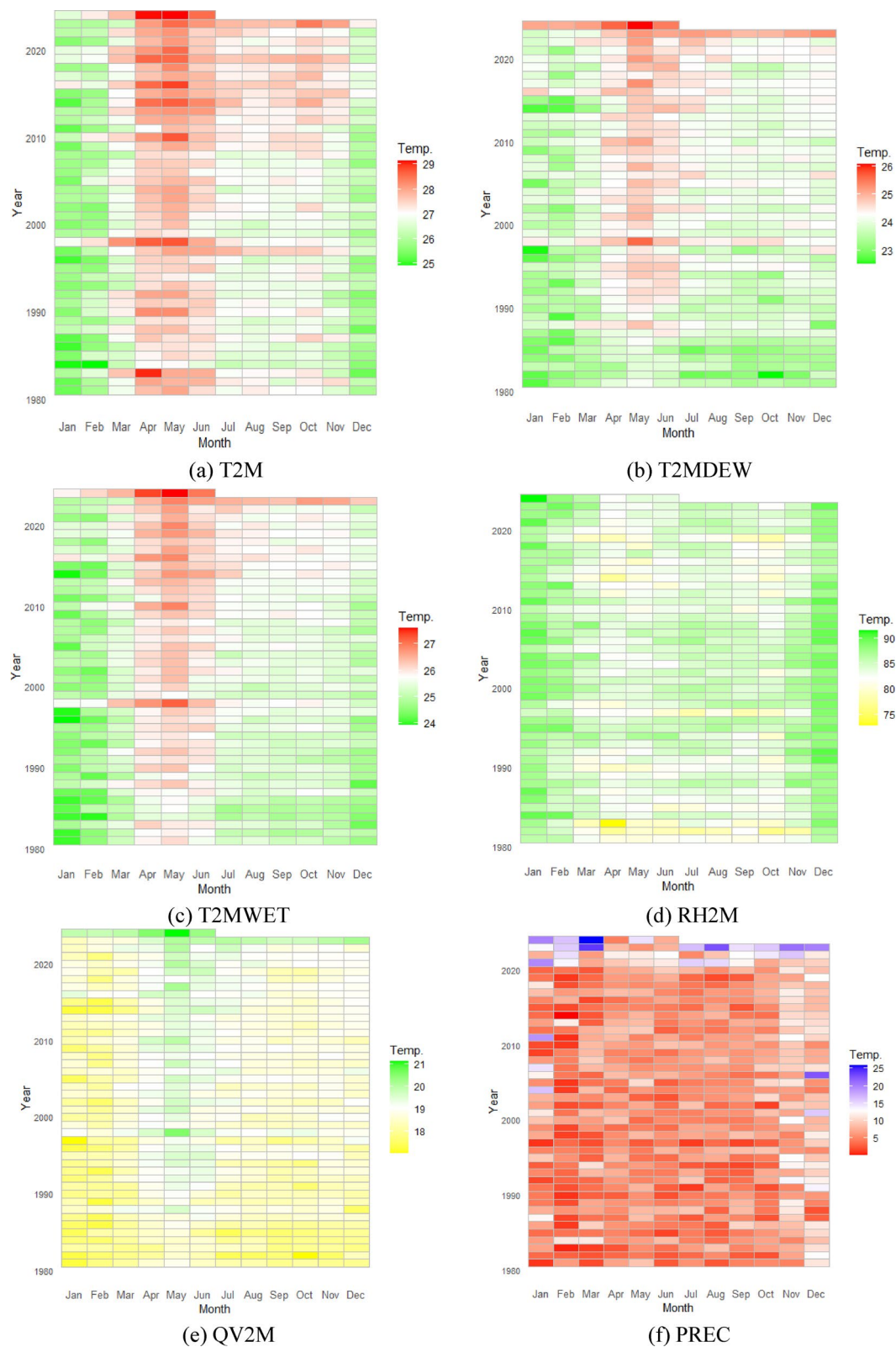
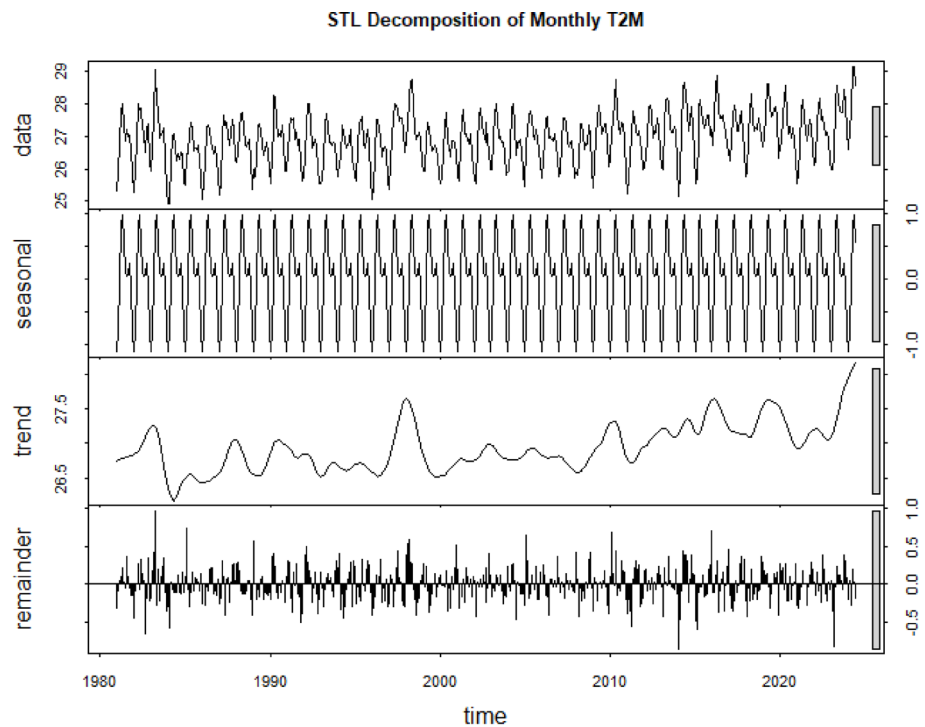
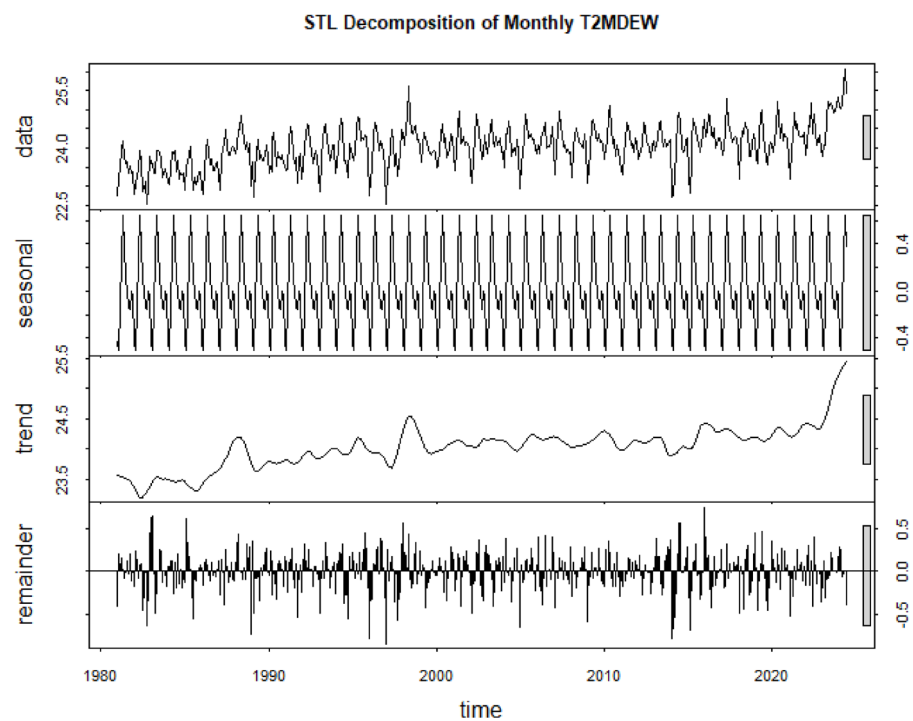


Fig. 3. Heatmaps of average monthly values for six climate variables in Johor Bahru from January 1981 to June 2024. Each panel shows a different variable: (a) T2M, (b) T2MDEW (temperature), (c) T2MWET, (d) RH2M, (e) QV2M, and (f) PREC.

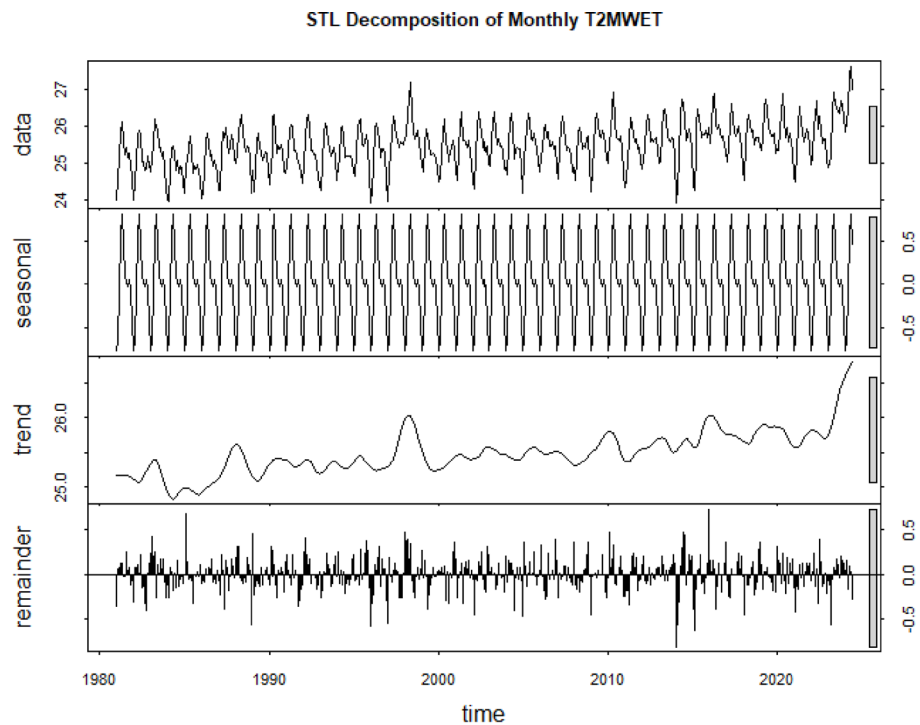


(a) STL decomposition for T2M

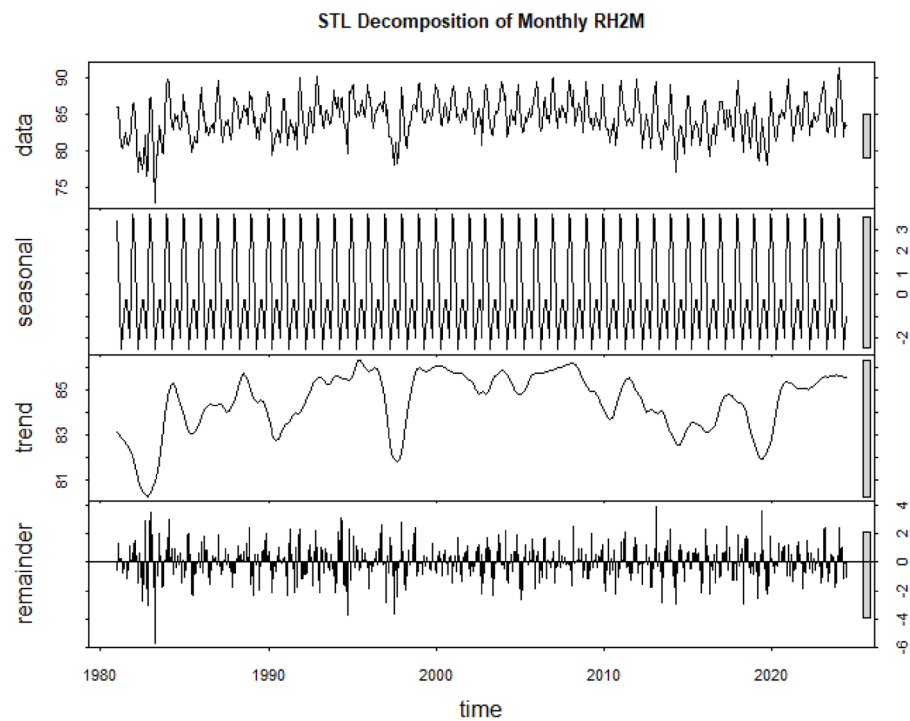


(b) STL decomposition for T2MDEW

Fig. 4. STL decomposition of monthly climate variables (1981–2024) for Johor Bahru. Each panel displays the decomposed components including trend, seasonal, and remainder, highlighting both long term changes and seasonal variation in the daily time series.



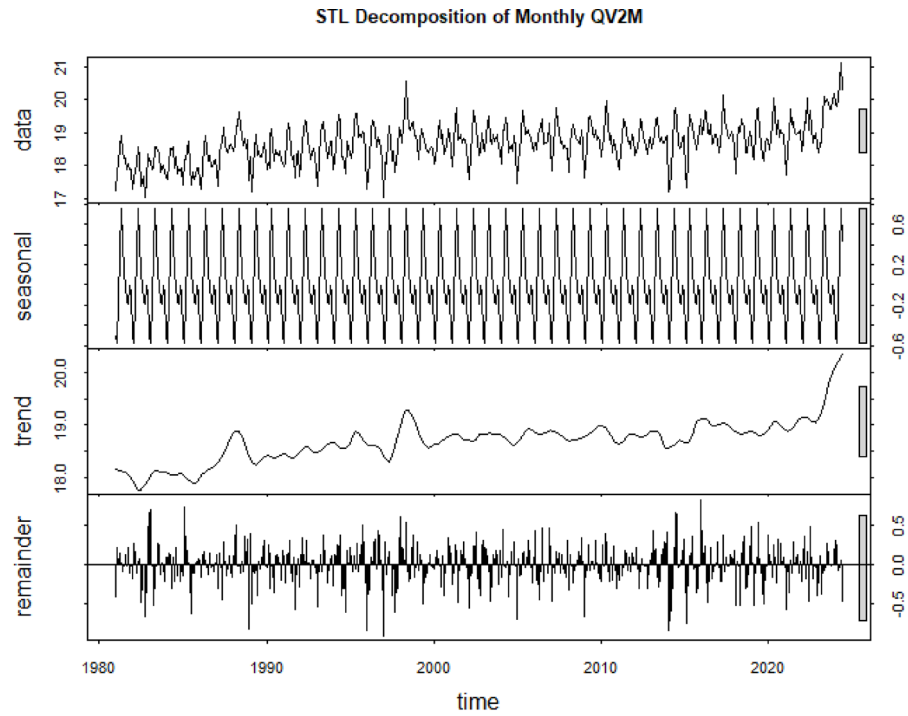
(c) STL decomposition for T2MWET



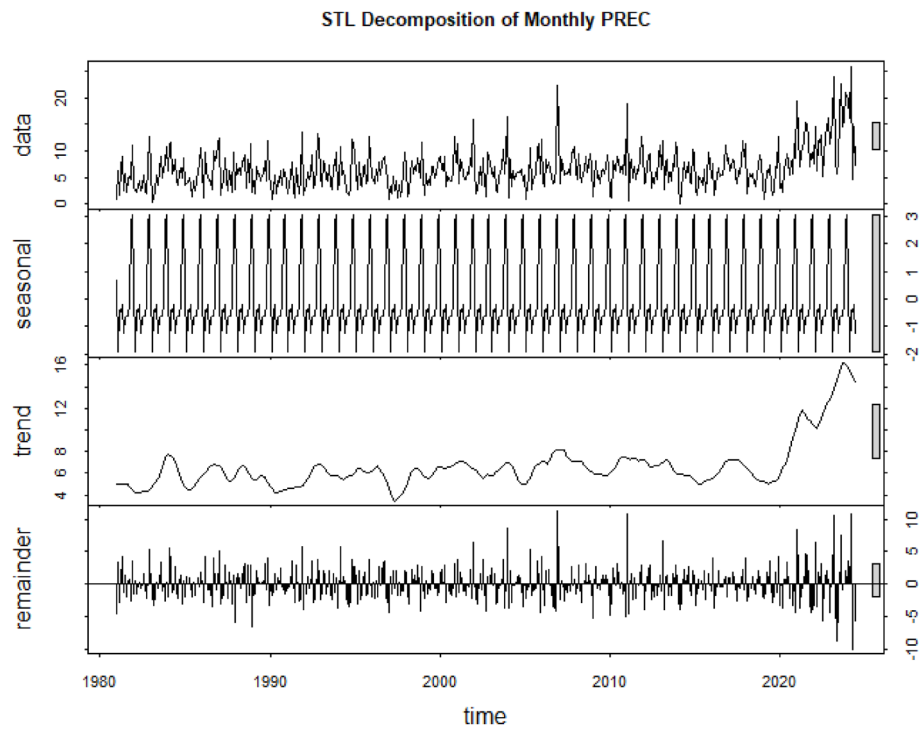
(d) STL decomposition for RH2M

Fig. 4. (continued)

reliability in out-of-sample forecasts. Meanwhile, Prophet underperformed on MAE, NSE, and KGE metrics, especially for PREC, where it recorded the lowest values across all metrics. This affirms Prophet's limitations in handling stochastic and weak seasonal climate variables.



(e) STL decomposition for QV2M



(f) STL decomposition for PREC

Fig. 4. (continued)

We display the visual analysis of NSE and KGE values recorded. Figures 7 and 8 depicted the NSE values for training and testing data respectively. Meanwhile, Figs. 9 and 10 display corresponding KGE values. These visual analysis reveals a consistent pattern. By examining closely, RF and SVR models outperform others in both metrics across most climate variables. RF shows particularly strong NSE and KGE scores in the training

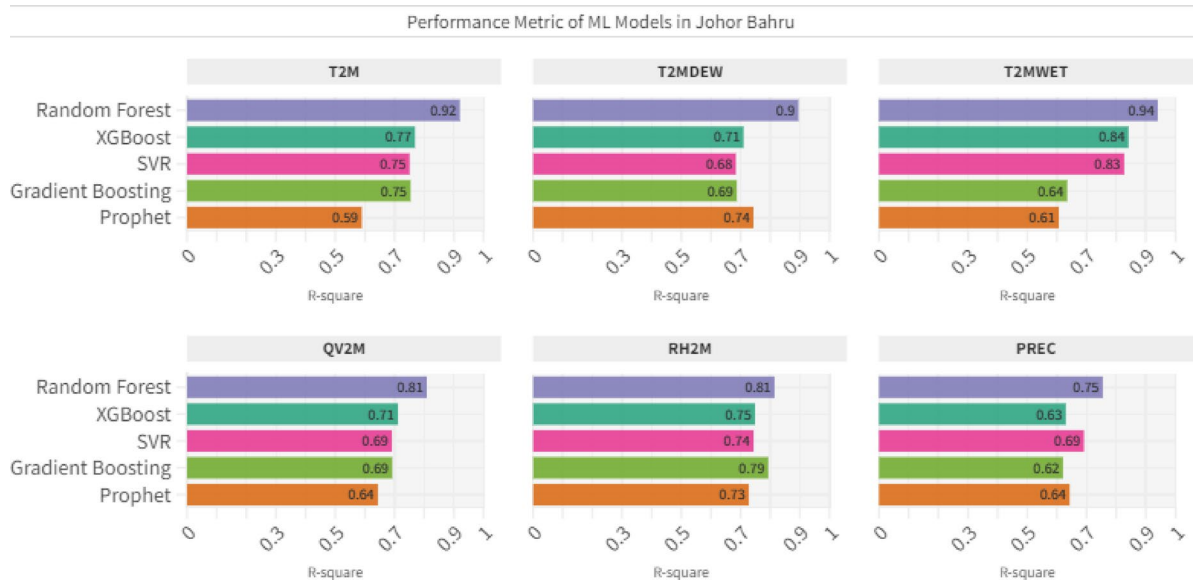


Fig. 5. R² performance metrics for training data across five machine learning models.

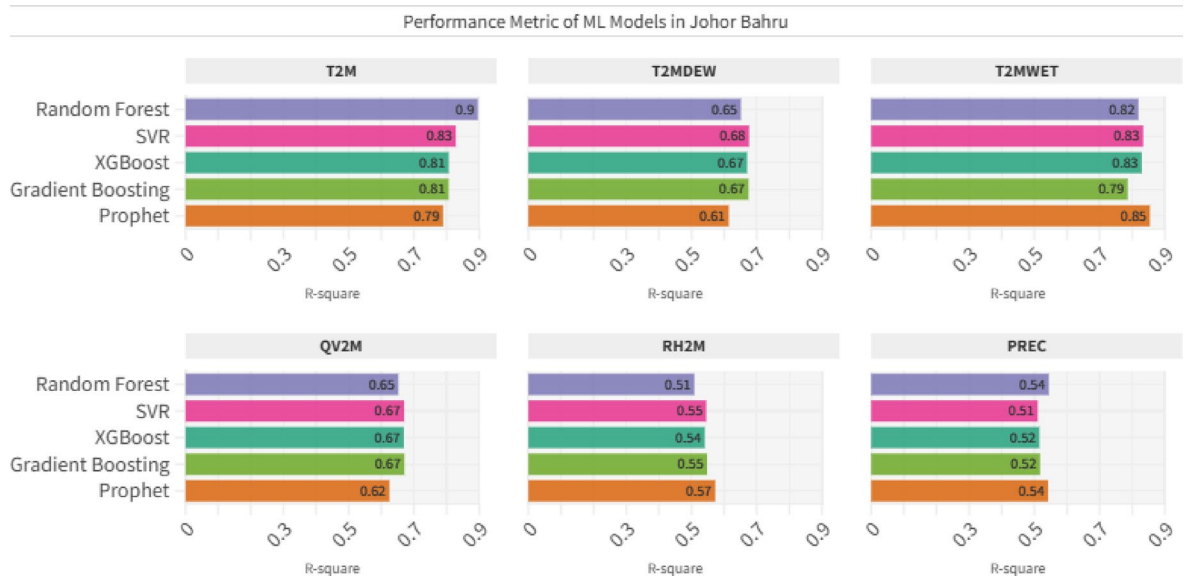


Fig. 6. R² performance metrics for testing data across five machine learning models.

phase, especially for temperature-related variables indicating high explanatory power and stability. In contrast, SVR demonstrates superior generalization in the testing phase, with the highest KGE values for most variables, confirming its reliability in out-of-sample forecasting. Prophet performs the weakest across both metrics, especially for PREC, reinforcing earlier findings of its limitations in modeling variables with weak seasonality and abrupt variability. These visualizations affirm the robustness of tree-based and kernel-based methods over decomposition-based models for daily climate prediction in Johor Bahru.

The Taylor diagrams obtained in Fig. 11 reveals several key insights. First, among the five ML models employed the SVR, RF, GBM and XGBoost demonstrated almost similar and strong performance as illustrated by close to proximity of the Taylor diagram except in Fig. 6(e). This suggests that these ML models are robust and reliable for climate variables prediction. However, the ML models employed exhibited significant difference patterns as shown in Fig. 6(e). The difference in behaviour among ML models is attributed to inferior performance due to its lower correlation and higher prediction error when predicting the precipitation in Johor Bahru. This finding is expected, since we have noticed the high values of RMSE recorded in Tables 3 and 4. Second, RF performed strongly, often rivaling SVR and XGBoost, especially in predicting humidity-related variables, though with slightly more variability in its predictions. Third, Prophet's performance was generally weaker since it is the farthest from the reference point in Fig. 11. The lower correlations values and higher errors highlight Prophet's

ML		T2M	T2MDEW	T2MWET	QV2M	RH2M	PREC
RF	RMSE	0.2182	0.2291	0.1621	0.3502	1.4444	5.4276
	MAE	0.1679	0.1750	0.1251	0.2701	1.1090	2.9374
	NSE	0.9396	0.8924	0.9450	0.8074	0.8543	0.7591
	KGE	0.9171	0.8516	0.9264	0.8074	0.7961	0.6122
XGBoost	RMSE	0.3671	0.3750	0.2615	0.4306	2.4303	8.7978
	MAE	0.2842	0.2883	0.2027	0.3316	1.881	5.0758
	NSE	0.8292	0.7119	0.8570	0.7088	0.5876	0.3670
	KGE	0.8660	0.7636	0.8868	0.7660	0.6545	0.3428
SVR	RMSE	0.3844	0.3951	0.2725	0.4440	2.5116	10.2959
	MAE	0.2955	0.2994	0.2086	0.3381	1.9204	4.7364
	NSE	0.8128	0.6801	0.8447	0.6904	0.5595	0.1332
	KGE	0.8674	0.7542	0.8865	0.7635	0.6567	0.0478
GBM	RMSE	0.3822	0.3937	0.27225	0.4447	2.4961	9.7114
	MAE	0.2952	0.3008	0.2098	0.3407	1.9297	5.4258
	NSE	0.8149	0.6823	0.84500	0.6893	0.5615	0.2074
	KGE	0.8608	0.7460	0.87936	0.7565	0.6371	0.2020
Prophet	RMSE	0.5299	0.5344	0.4241	0.6050	2.9858	10.7757
	MAE	0.4140	0.4076	0.3246	0.4648	2.3546	5.3306
	NSE	0.6477	0.4203	0.6257	0.4325	0.3829	0.0479
	KGE	0.7199	0.4986	0.7019	0.5121	0.4513	0.2633

Table 3. Performance metrics for training data across five machine learning models based on four evaluation indicators.

ML		T2M	T2MDEW	T2MWET	QV2M	RH2M	PREC
RF	RMSE	0.4039	0.4105	0.2926	0.4645	2.6538	9.7572
	MAE	0.3145	0.3158	0.2232	0.3585	2.0330	5.6366
	NSE	0.7955	0.6515	0.8190	0.6521	0.5086	0.1441
	KGE	0.8591	0.7474	0.8741	0.7355	0.6301	0.1715
XGBoost	RMSE	0.3929	0.3993	0.2845	0.4532	2.5659	9.6920
	MAE	0.3034	0.3048	0.2170	0.3489	1.9707	5.4925
	NSE	0.8065	0.6703	0.8289	0.6688	0.5406	0.1555
	KGE	0.8608	0.7571	0.8786	0.7509	0.6347	0.1534
SVR	RMSE	0.3916	0.3958	0.2811	0.4525	2.5529	9.9720
	MAE	0.3018	0.3013	0.2123	0.3461	1.9416	4.8563
	NSE	0.8078	0.6760	0.8330	0.6698	0.5453	0.1060
	KGE	0.8685	0.7617	0.8851	0.7539	0.6537	0.0237
GBM	RMSE	0.3931	0.3967	0.2845	0.4521	2.5458	9.5835
	MAE	0.3035	0.3022	0.2164	0.3477	1.9484	5.5080
	NSE	0.8066	0.6745	0.8288	0.6704	0.5484	0.1761
	KGE	0.8601	0.7551	0.8754	0.7490	0.6350	0.1643
Prophet	RMSE	1.2290	1.2333	1.0994	1.4194	4.9750	13.7079
	MAE	0.9865	1.0356	0.9330	1.2245	3.8790	9.2601
	NSE	0.9155	0.7861	0.6896	0.2721	0.2728	0.2332
	KGE	0.8813	0.1093	0.2895	0.2987	0.2702	0.2791

Table 4. Performance metrics for testing data across five machine learning models four evaluation indicators.

limitations in these scenarios illustrated by Taylor diagram. These findings may be a direct consequence of failing to handle the dry seasons properly in Johor Bahru. Finally, we recognized RF achieves the lowest RMSE as shown in Table 3 across all climate variables on the training dataset, indicating a strong in-sample fit. On the other hand, Table 4 reveals that SVR consistently attains the lowest RMSE on the testing dataset for five out of six variables. These results suggest that while RF captures training patterns effectively, SVR demonstrates superior generalization performance in out-of-sample forecasting.

To assess whether differences in model performance were statistically significant, we conducted a Friedman test on RMSE values across the six climate variables (T2M, T2MDEW, T2MWET, QV2M, RH2M, and PREC).

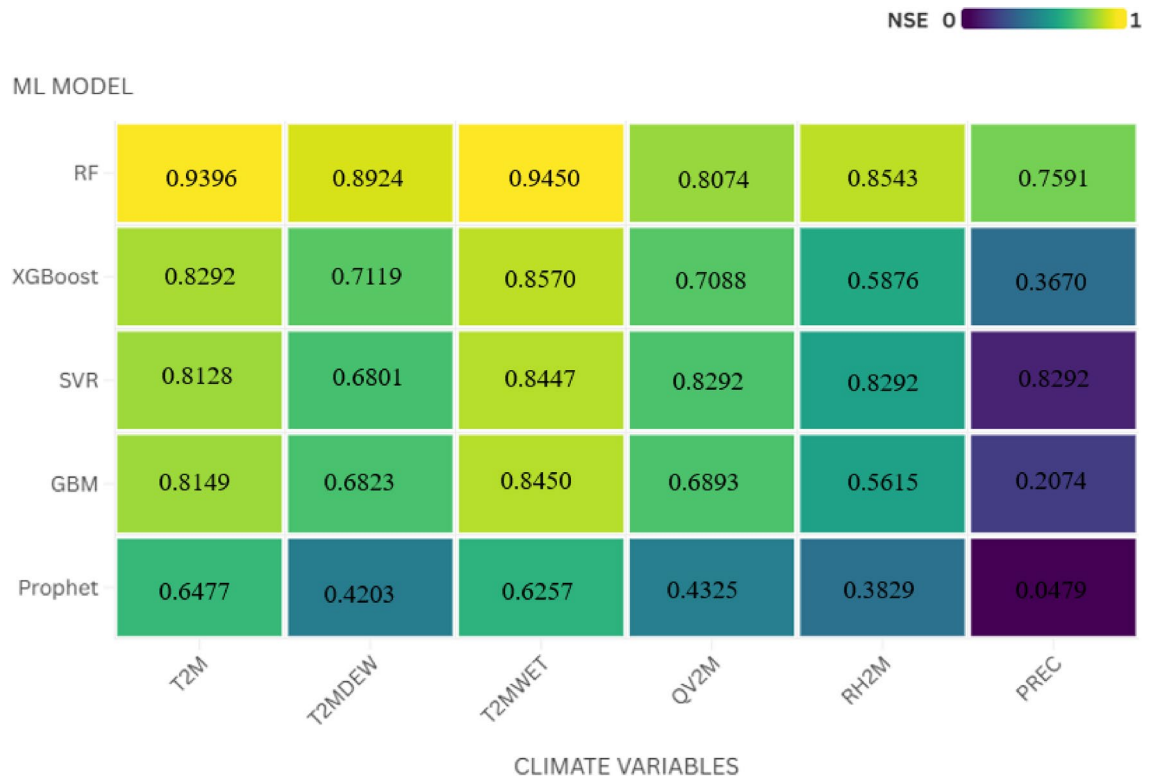


Fig. 7. Comparison of NSE for training data across five machine learning models.

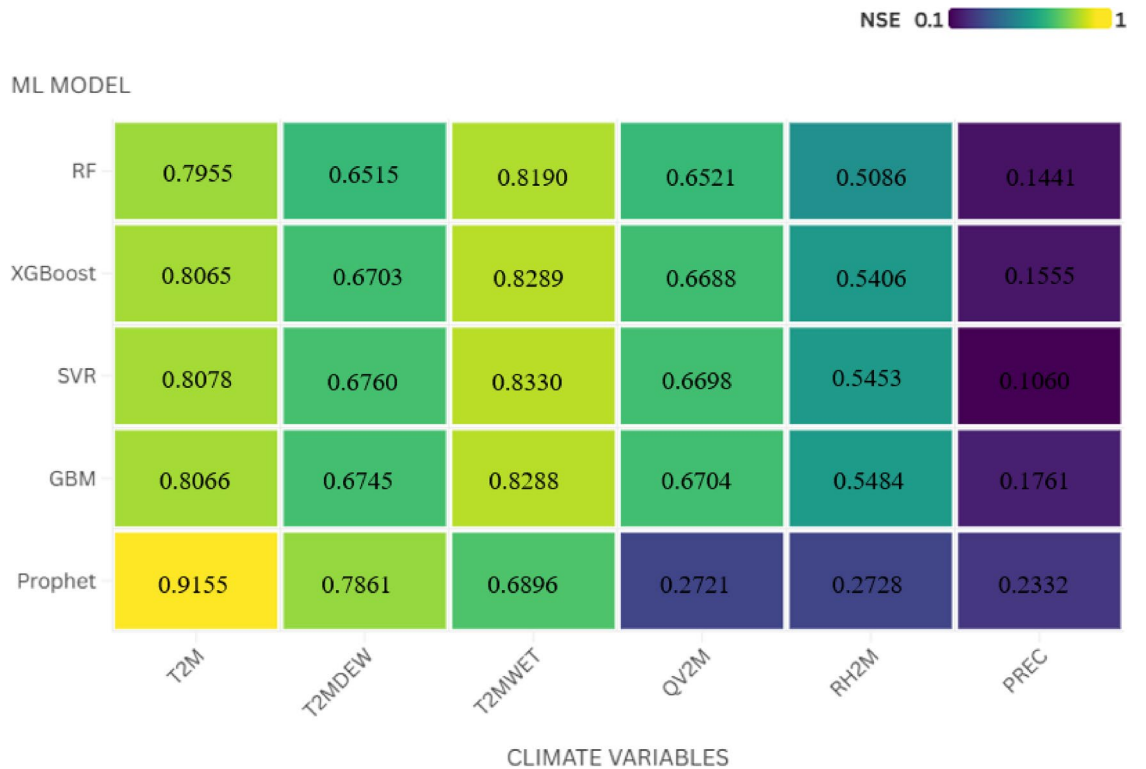


Fig. 8. Comparison of NSE for testing data across five machine learning models.

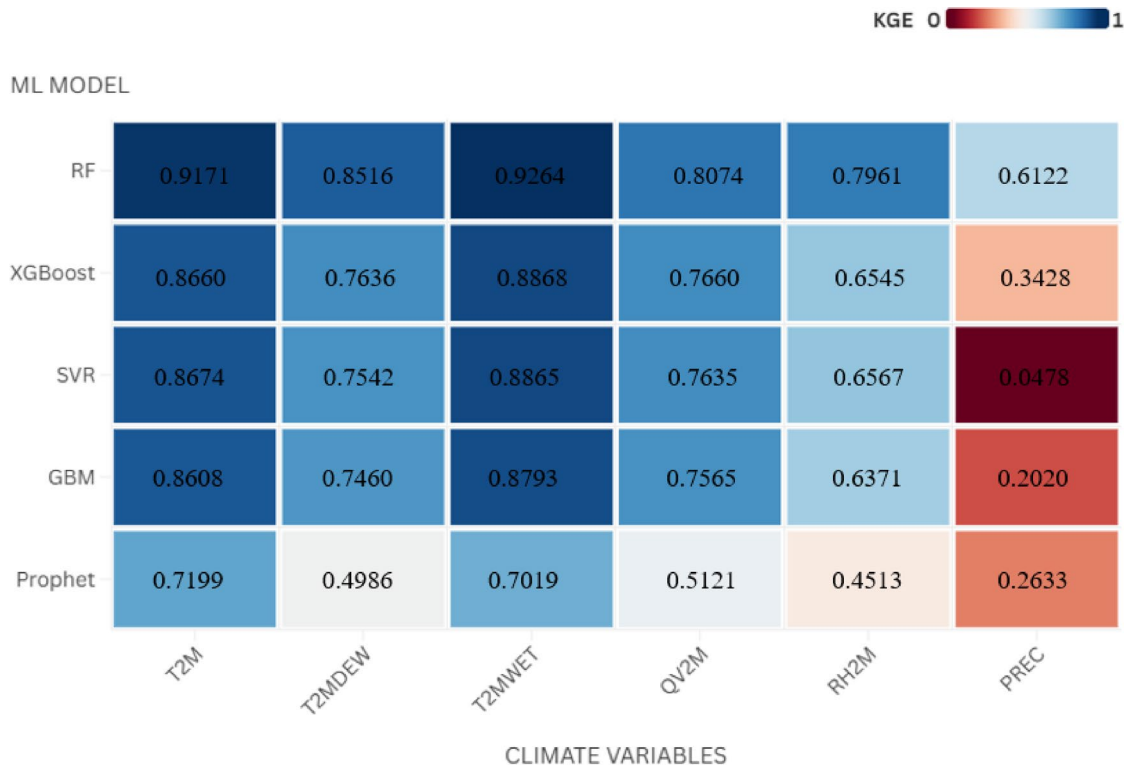


Fig. 9. Comparison of KGE for training data across five machine learning models.

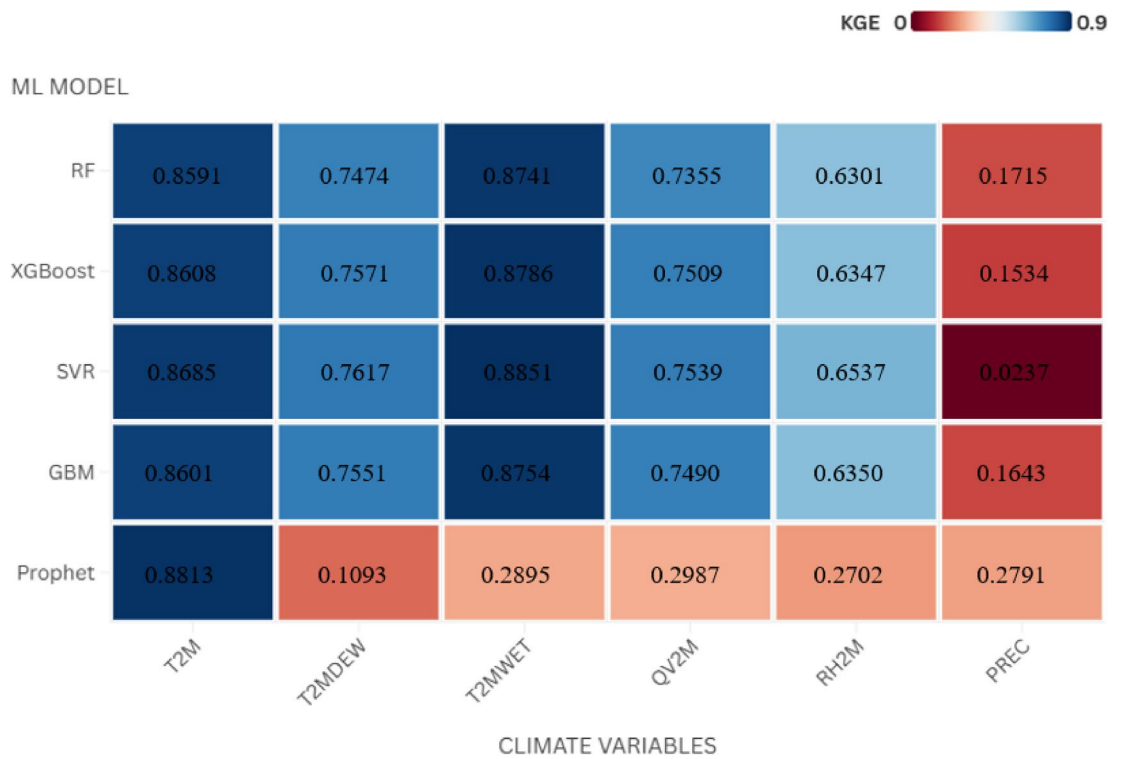


Fig. 10. Comparison of KGE for testing data across five machine learning models.

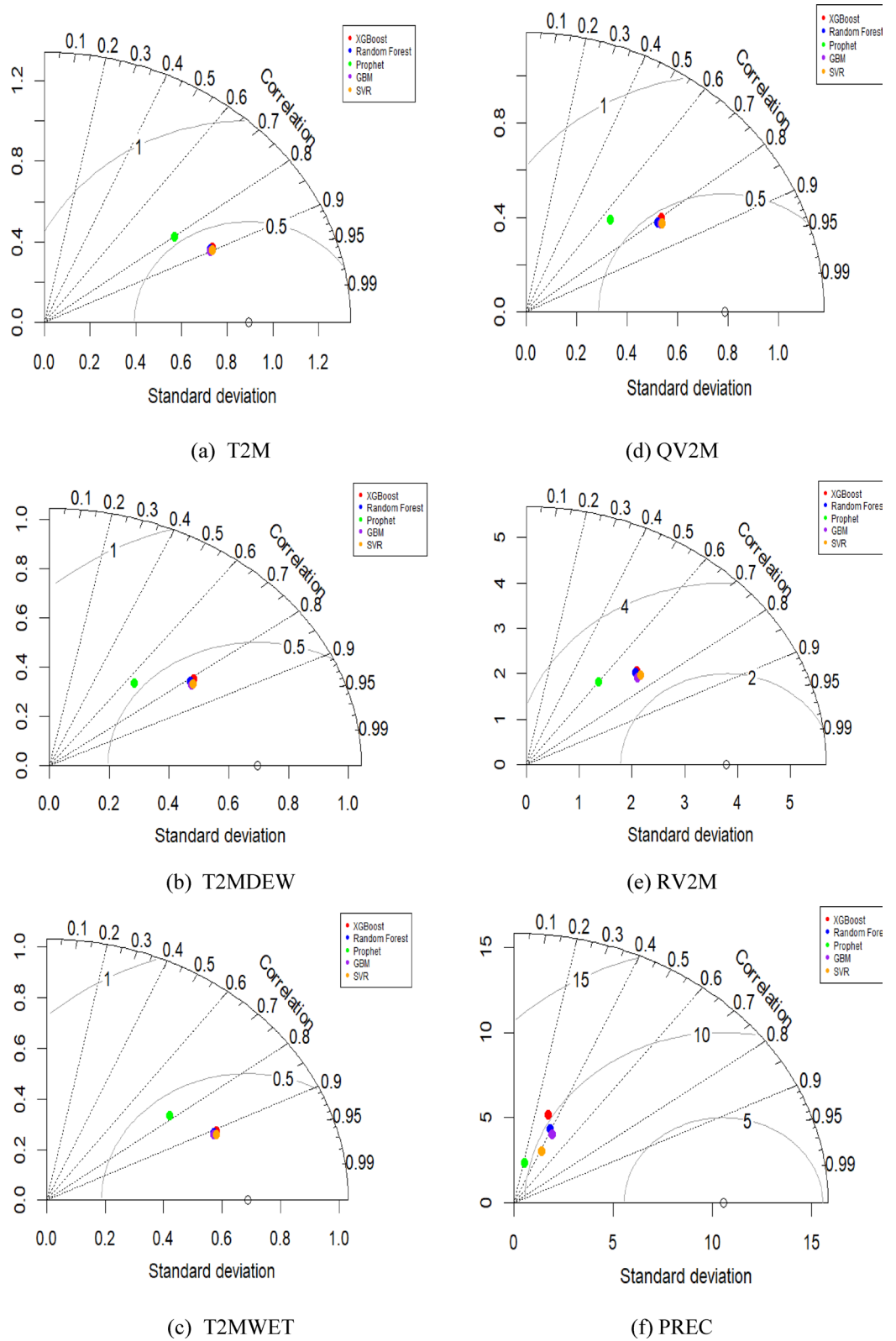


Fig. 11. Performance comparisons of five models using Taylor diagram. The diagram simultaneously displays correlation, standard deviation, and centered root mean square error to assess the models' accuracy and consistency in predicting climate variables.

Model 1	Model 2	Rank difference	Significance difference
XGBoost	Prophet	2.5	Yes
XGBoost	RF	1.17	No
XGBoost	SVR	0.67	No
XGBoost	GBM	0.50	No
Prophet	RF	1.33	No
Prophet	SVR	3.17	Yes
Prophet	GBM	3.00	Yes
RF	SVR	1.83	No
RF	GBM	1.67	No
SVR	GBM	0.17	No

Table 5. Nemenyi post-hoc test results for pairwise model comparison based on RMSE rank differences.

Model	Training time (Relative)	Computational demand	Prediction accuracy (RMSE)	Notable characteristics
RF	Moderate	Moderate	High	May overfit, robust to noise
SVR	Low	Low	High	Fast in training, sensitive to outliers
XGBoost	High	High	High	Best overall accuracy
GBM	Moderate high	Moderate high	High	Good accuracy, sensitive to hyperparameter
Prophet	Low	Low	Low	Fastest in computational speed, underperform on weakly seasonal component

Table 6. Summary of computational characteristics and prediction accuracy of ML models for climate prediction in Johor bahru.

The Friedman's test yielded a statistically significant result ($\chi^2 = 23.63, p = 0.00026$), indicating that the observed differences in model performance were not due to random variation. A post-hoc Nemenyi test was performed to determine pairwise differences between models. The critical difference (CD) was calculated to be 2.49. Table 5 summarizes the rank differences and indicates whether the differences are statistically significant. These results confirm that SVR, GBM, and XGBoost significantly outperform Prophet, while differences among the top-performing models (SVR, RF, GBM, and XGBoost) are not statistically significant. This supports our conclusion that SVR and RF are highly suitable for climate variable prediction in Johor Bahru, while Prophet's performance is statistically inferior for the studied context.

To complement our evaluation of model performance based on performance metrics, we qualitatively assessed the computational efficiency and training characteristics of each machine learning model, as summarized in Table 6. We compared the models not only in terms of accuracy but also their computational efficiency. SVR and Prophet were the fastest to train and used the least resources. SVR maintained high prediction accuracy, while Prophet showed weaker results, especially for precipitation data. RF and GBM took more time to train but performed well in prediction. XGBoost needed the most time and computing power but gave the best overall accuracy. These differences highlight the need to choose models based on both performance and available computing resources, especially in settings where speed or hardware limitations are important.

Discussion

The analysis of climate variables in Johor Bahru reveals several key insights about the ML models employed in terms of performance metrics used. First, the results suggest that for climate prediction in Johor Bahru, SVR and RF should be prioritized due to their consistent performance across various climate variables related to temperature. These models' ability to handle the complexities and variability inherent in climate data makes them invaluable tools for improving regional climate forecasts and informing adaptive strategies in agriculture, water management⁸ and other climate-sensitive sectors³⁷. While GBM, XGBoost and Prophet have their merits, particularly in simpler or more linear data contexts, SVR stands out as the most reliable and versatile models for this study. Its application can lead to more accurate and actionable climate predictions, ultimately supporting better decision-making and resource management in the face of ongoing climate change⁶.

Second, Figs. 5 and 6 further illustrate a noteworthy contrast in model behavior. RF also performed strongly, often rivaling XGBoost, especially in predicting humidity-related variables. This indicates its robustness and versatility in handling different types of climate data⁸. RF achieves the highest R^2 in training, reflecting a strong ability to capture in-sample variation. However, SVR records the highest R^2 in the testing set for most variables, suggesting superior generalization performance. Its ability to achieve high R^2 values suggests that it is highly effective in capturing the underlying trends and variability in climate data²³. This pattern is consistent with RMSE-based evaluations in Tables 3 and 4 and highlights the importance of evaluating models not just on fit but on predictive robustness. Therefore, while RF excels in modeling complex patterns during training, SVR demonstrates better adaptability in forecasting new, unseen climate conditions.

Third, Prophet's performance was generally weaker where non-linear and complex climate patterns are prevalent. The lower R^2 values highlight Prophet's limitations in these scenarios. The STL decomposition of

the PREC variable revealed a weakly defined seasonal component and significant irregular fluctuations. These characteristics pose challenges for models like Prophet, which rely on strong seasonal patterns and smooth trends for accurate forecasting. The model's additive structure and assumptions about seasonality make it less suitable for capturing the abrupt changes and stochastic variability observed in precipitation data. As noted by³⁸ this structural assumption leads to flat and dampened predictions that are poorly suited to capturing the irregular and abrupt nature of real-world climate extremes. Prophet, while useful in more straightforward, linear scenarios, may require enhancements or hybridization with other models to improve its performance in complex environments.

Fourth, the results of this study show that the machine learning models applied, particularly RF, GBM, and XGBoost, demonstrate strong performance in predicting daily precipitation and humidity. These variables are essential for anticipating flood events, especially in Johor Bahru, where rapid urban development and heavy rainfall have made flash floods increasingly common. The ability to generate accurate short-term forecasts provides a valuable foundation for enhancing flood early warning systems. By incorporating these models into real-time monitoring and alert mechanisms, local authorities could better anticipate high-risk conditions, issue timely warnings, and implement responsive measures to reduce flood-related impacts. To improve future flood risk prediction and support operational early warning systems, it would be beneficial to combine climate-based machine learning models with hydrological or groundwater modeling frameworks. Such integration would provide a more complete picture of flood processes and enhance the reliability of alerts for both decision-makers and affected communities. This approach contributes directly to sustainable urban management and aligns with Sustainable Development Goal 11, which emphasizes the importance of building resilient and sustainable cities that are prepared to manage environmental risks and climate-related disasters.

Finally, although this study did not use future climate projections, several steps were taken to support the robustness of the predictions under changing climate conditions. The dataset includes more than 40 years of daily data from 1981 to 2024, covering a wide range of seasonal patterns and extreme weather events. This helps the models learn from long-term trends and variability. We also used different machine learning models to compare their performance and identify those that generalize better across time. In addition, the models were tested on a separate portion of the data that was not used for training, to ensure the results reflect realistic forecasting performance. We acknowledge that future studies could include projected climate data or scenario-based modeling to better assess long-term changes.

Conclusion

This study presents an in-depth analysis of climate trends and evaluates the performance of various machine learning models in predicting key climate variables in Johor Bahru. The heatmap analysis suggests that both temperature and humidity related variables show significant upward trends since 1981. These trends indicate an increased risk of heat stress, which could have serious impacts on public health and agricultural productivity in the region. Among the machine learning models evaluated, SVR and RF consistently delivered superior prediction accuracy, particularly for temperature-related variables. These models demonstrated strong predictive capabilities, making them valuable tools for future climate prediction efforts. The results suggest that these models should be given priority in climate prediction efforts, particularly in regions facing similar climatic challenges. Accurate climate predictions are crucial in the context of global climate change, especially for sectors such as agriculture that are highly vulnerable to climate variability. The use of advanced machine learning models in this area offers a promising approach to strengthening the resilience of these sectors. The findings of this study have significant implications for climate adaptation strategies in Johor Bahru and similar tropical urban settings. Specifically, the daily forecasts generated by the models can be used by policymakers for operational decision-making. These include activating flood warnings based on projected rainfall thresholds, issuing public health advisories during periods of heat stress, revising urban infrastructure design to account for increasing rainfall intensity, and supporting investment decisions in climate-resilient urban systems. Future research should aim to further refine these models and explore their application in different climate regions to improve prediction accuracy and support adaptation strategies in response to ongoing climate change. The insights from this research will be critical in shaping policies and investment decisions, thereby strengthening society's resilience in the face of ongoing climate change.

Data availability

Source of data is provided within the manuscript which can be access freely.

Received: 19 February 2025; Accepted: 18 June 2025

Published online: 02 July 2025

References

- Zheng, Y. et al. Assessing the impacts of climate variables on long-term air quality trends in Peninsular Malaysia. *Science Total Environment* **901**, 166430 (2023).
- Huang, L. et al. Solar radiation prediction using different machine learning algorithms and implications for extreme climate events. *Front Earth Sci. (Lausanne)* **9**, 596860 (2021).
- Ge, J. et al. Prediction of Greenhouse Tomato Crop Evapotranspiration Using XGBoost Machine Learning Model. *Plants* **11**, (2022).
- Pathan, A. I. et al. Comparative assessment of rainfall-based water level prediction using machine learning (ML) techniques. *Ain Shams Eng. J.* **15**, 102854 (2024).
- Dias Júnior, D. A. et al. Automatic method for classifying COVID-19 patients based on chest X-ray images, using deep features and PSO-optimized XGBoost. *Expert Syst. Appl.* **183**, 115452 (2021).

6. Bhoopathi, S., Kumar, N. & Pal, M. Somesh Evaluating the performances of SVR and XGBoost for short-range forecasting of heatwaves across different temperature zones of India. *Applied Comput. Geosciences* **24**, 100204 (2024).
7. Satpathi, A. et al. Evaluating statistical and machine learning techniques for sugarcane yield forecasting in the Tarai region of North India. *Comput. Electron. Agric.* **229**, 109667 (2025).
8. Bargam, B. et al. Evaluation of the support vector regression (SVR) and the random forest (RF) models accuracy for streamflow prediction under a data-scarce basin in Morocco. *Discover Appl. Sciences* **6**, 306 (2024).
9. Bhoopathi, S., Sumanth, K., Akanksha, L. & Pal, M. Evaluating the performance of ANN, SVR, RF, and XGBoost in the prediction of maximum temperature and heat wave days over rajasthan, India. *J Hydrol. Eng* **29**, 6243 (2024).
10. Fang, Z. G., Yang, S. Q., Lv, C. X., An, S. Y. & Wu, W. Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study. *BMJ Open* **12**, 7 (2022).
11. Cesarini, L. et al. Comparison of deep learning models for milk production forecasting at National scale. *Comput. Electron. Agric.* **221**, 108933 (2024).
12. Chen, T., Guestrin, C. & XGBoost A scalable tree boosting system. in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* vols 13-17-August-2016 785–794 Association for Computing Machinery, (2016).
13. Brkić, Ž. & Larva, O. Impact of climate change on the Vrana lake surface water temperature in Croatia using support vector regression. *J. Hydrol. Reg. Stud.* **54**, 101858 (2024).
14. Miftahurrohmah, B., Kuswanto, H., Pambudi, D. S., Fauzi, F. & Atmaja, F. Assessment of the support vector regression and random forest algorithms in the Bias correction process on temperatures. *Procedia Comput. Sci.* **234**, 637–644 (2024).
15. Velasco, L. C., Estose, A. J., Opon, M., Tabanao, E. & Apdian, F. Performance evaluation of support vector regression machine models in water level forecasting. *Procedia Comput. Sci.* **234**, 436–447 (2024).
16. Fronzi, D. et al. Towards Groundwater-Level prediction using prophet forecasting method by exploiting a High-Resolution hydrogeological monitoring system. *Water (Switzerland)* **16**, 152 (2024).
17. Elneel, L., Zitouni, M. S., Mukhtar, H. & Al-Ahmad, H. Examining sea levels forecasting using autoregressive and prophet models. *Sci Rep* **14**, 14337 (2024).
18. Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I. & Matsopoulos, G. K. A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet* vol. 15 Preprint at (2023). <https://doi.org/10.3390/fi15080255>
19. Kisi, O., Tombul, M. & Kermani, M. Z. Modeling soil temperatures at different depths by using three different neural computing techniques. *Theor. Appl. Climatol.* **121**, 377–387 (2015).
20. Latif, S. D. et al. Prediction of atmospheric carbon monoxide concentration utilizing different machine learning algorithms: A case study in Kuala Lumpur, Malaysia. *Environ. Technol. Innov.* **32**, 103387 (2023).
21. Zakaria, M. N. A. et al. Exploring machine learning algorithms for accurate water level forecasting in Muda river. *Malaysia Heliyon.* **9**, e17689 (2023).
22. Asif, M., Kuglitsch, M. M., Pelivan, I. & Albano, R. Review and intercomparison of machine learning applications for Short-term flood forecasting. *Water Resour. Manage.* <https://doi.org/10.1007/s11269-025-04093-x> (2025).
23. Alawi, O. A., Kamar, H. M., Homod, R. Z. & Yaseen, Z. M. Deep learning and tree-based models for Earth skin temperature forecasting in Malaysian environments. *Appl. Soft Comput.* **155**, 111411 (2024).
24. Md Ali, Z., Wai Tan, L. & Masyra Daud, I. & Aliza ahmad, N. *Rainfall characteristics of Johor Bahru and Kota bharu, Malaysia. Journal Sci. Technology* **9**, 77–83 (2017).
25. Paudel, D. et al. Machine learning for large-scale crop yield forecasting. *Agric. Syst.* **187**, 103016 (2021).
26. Muhammad, M. K. I. et al. Heatwaves in Peninsular Malaysia: a Spatiotemporal analysis. *Sci Rep* **14**, 4255 (2024).
27. Chuah, C. J., Ho, B. H. & Chow, W. T. L. Trans-boundary variations of urban drought vulnerability and its impact on water resource management in Singapore and Johor, Malaysia. *Environmental Res. Letters* **13**, 074011 (2018).
28. Taylor, S. J. & Letham, B. Forecasting at scale. *Am. Stat.* **72**, 37–45 (2018).
29. Cortes, C., Vapnik, V. & Saïtta, L. *Support-Vector networks editor. Machine Learning* **20**, 273–297 (1995).
30. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. Preprint at (1999). <http://www.salford-systems.com/treenet.html>
31. Breiman, L. *Random Forests*. vol. 45 (2001).
32. Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J. Hydrol. (Amst)*. **377**, 80–91 (2009).
33. Muhammad, M. K. I. et al. The development of evolutionary computing model for simulating reference evapotranspiration over Peninsular Malaysia. *Theor. Appl. Climatol.* **144**, 1419–1434 (2021).
34. Abdul Talib, S. A., Idris, W. M. R., Neng, L. J. & Lihan, T. Abdul rasid, M. Z. Irregularity and time series trend analysis of rainfall in Johor. *Malaysia Heliyon.* **10**, e30324 (2024).
35. Tan, M. L., Chua, V. P., Li, C. & Brindha, K. Spatiotemporal analysis of hydro-meteorological drought in the Johor river basin, Malaysia. *Theor. Appl. Climatol.* **135**, 825–837 (2019).
36. Bong, C. H. J. & Richard, J. Drought and climate change assessment using standardized precipitation index (Spi) for Sarawak river basin. *J. Water Clim. Change.* **11**, 956–965 (2020).
37. Zennaro, F. et al. Exploring machine learning potential for climate change risk assessment. *Earth Sci. Rev* **220**, 103752 (2021).
38. Li, D. et al. Prediction of rainfall time series using the hybrid DWT-SVR-Prophet model. *Water (Switzerland)* **15**, 1935 (2023).

Acknowledgements

This work was supported by Universiti Putra Malaysia (Grant number GP-IPM-9773500).

Author contributions

F.Z.C. R. was responsible for the conceptualization, methodology, software development, data curation, formal analysis, original draft preparation, and funding acquisition. N.A.K.R. contributed to visualization, investigation, and the review and editing of the manuscript. M.F.M. provided supervision, validation, and contributed to the review and editing process. All authors have read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.Z.C.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025