Journal of Informatics and Web Engineering
Vol. 4 No. 1 (February 2025) eISSN: 2821-370X

Climate Change Analysis in Malaysia Using Machine Learning

Anishalache Subramanian[1], Naveen Palanichamy[2,*], Kok-Why Ng[3], Sandhya Aneja[4]
[1,2,3]Faculty of Computing and Informatics Multimedia University, Persiaran Multimedia, 63100, Cyberjaya, Malaysia.
[4]School of Computer Science and Mathematics, Marist College, Poughkeepsie, NY12601, United States.
**corresponding author: (p.naveen@mmu.edu.my; ORCiD: 0000-0003-4601-9770)*

Abstract - Climate change presents significant challenges to ecosystems, economies, and societies globally. In Malaysia, a tropical country highly dependent on its natural resources, the impacts are evident in altered rainfall patterns, rising temperatures, and extreme weather events. Despite these challenges, many studies still predominantly rely on traditional statistical methods, which limit their capacity for making accurate climate predictions and developing effective policy solutions.This study effectively addresses the existing gap in research by analyzing extensive historical climate data using advanced machine learning (ML) techniques. The primary focus is on accurately forecasting trends in both precipitation patterns and surface air temperature fluctuations. Performance measures like Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are used to assess three ML models: Support Vector Regression (SVR), Random Forest Regression (RFR) and Linear Regression (LR). The findings demonstrate that LR performs better than the other models in forecasting patterns of precipitation and temperature. The results suggest a significant increase in temperature and unpredictable patterns of precipitation, and that poses major implications for agriculture, infrastructure resilience, and water management. Malaysia's climate resilience is improved by this research, which promotes data-driven policymaking by assessing current climate adaptation methods and offering practical ideas.

Keywords—Machine Learning, Temperature , Precipitation, Support Vector Regression, Random Forest, Linear Regression

Received: 01 October 2024; Accepted: 17 December 2024; Published: 16 February 2025

1. INTRODUCTION

Climate change presents huge challenges for Malaysia, a tropical country whose economy is highly dependent on natural resources. Climate change affects the rise in temperature and irregular patterns of precipitation that impact agriculture, public health, and infrastructure. Increases in temperature cause increased incidences of heatwaves, reduced agricultural productivity, and enhanced vulnerability to health risks. Meanwhile, erratic rainfall increases the intensities of urban flooding, which causes infrastructural damages and complicates the works associated with urban planning. The increasing occurrence of such extreme weather phenomena as floods and droughts in Malaysia poses a huge risk to ecosystems and critical services like the management of water and energy resources [1].

ML techniques may be an encouraging approach in improving climate forecast accuracy [2], [3], [4], [5]. ML can deal with huge and complex data sets, and it could discover patterns that may be missed by traditional statistical methods.

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2025.4.1.22

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

The benefits of ML in discovering and revealing patterns from data do, however, come with challenges such as overfitting and dependency on good-quality data. Prior research has uncovered a pattern linking rises in temperature with falls in agricultural production levels for Malaysia, which has particularly focused on basic crops like rice. Another issue that arises from changes in rain fall patterns is a greater potential for flooding.

This study builds on such findings by using advanced ML approaches to analyze temperature and precipitation data, which, therefore, bring insights that could help develop effective climate adaptation strategies [8], [9]. In this respect, determining the best performing model in climate prediction will give policymakers and urban planners an opportunity to improve their management of the risks related to climate change[10], [11], [12]. This study evaluates three different ML models: SVR, RFR, and LR to predict future temperature and precipitation changes in Malaysia. Based on this regard, using some performance measures like MSE and RMSE, this paper seeks to determine a model with the highest degree of accuracy for its forecasts to directly rival the intrinsic failures associated with conventional methodologies [6], [7].

The organization of this paper is delineated as follows: It first commences with a literature review that situates the importance of ML within the context of climate prediction. Then, there is a large methodology section describing the processes of data collection and analytical techniques used. Further, the results section carries out a comparative evaluation of model performance with respect to the defined metrics. Finally, it discusses the results in relation to the existing literature and highlights their implications for decision-makers and urban developers in addressing climate-induced hazards.

2. LITERATURE REVIEW

In recent years, ML methodologies have emerged as a critical tool in the investigation of climate change, particularly in regions like Malaysia [13], [14], [15], [16]. The complex nature of climate data encompassing diverse meteorological, geographical, and temporal parameters—necessitates advanced analytical techniques. ML enables researchers to scrutinize these intricate data points and discern underlying patterns associated with climate trends and anomalies.

2.1. Machine Learning and Deep Learning Approaches

Recent studies emphasize the potential of ML and deep learning (DL) models in improving climate predictions. Table 1 shows the ML and DL techniques used. Convolutional neural networks (CNNs) paired with support vector machines (SVMs) have advanced precipitation and temperature forecasting by converting raw climate data into structured formats, leading to improved accuracy, especially in Malaysian datasets [15], [16]. Ensemble methods, like Random Forest (RF), have consistently outperformed Lasso regression in classifying severe weather, proving effective for complex climate data [6], [7].

Multimodal data fusion, combining satellite imagery with ground-based meteorological data, has enhanced the detection of climate anomalies. A deep forest model that incorporates factors such as humidity, wind speed, and solar radiation achieved an 85% precision rate, surpassing traditional classifiers like SVM and K-nearest neighbors (KNN) [8]. Additionally, stratified normalization has improved cross-regional climate predictions, addressing inter-regional data discrepancies.

In assessing climate impacts on ecosystems, a comparison between Random Forest and feed-forward neural network (FFNN) classifiers showed FFNN achieving 78% accuracy, underscoring the value of DL in evaluating environmental changes [5]. These advancements in ML offer more accurate tools for climate forecasting and impact assessment.

2.2. Evaluation Metrics in Climate Prediction Models

Recent studies have used diverse evaluation metrics to assess ML models for climate prediction. [9], [17] adopted metrics such as Precision, MAE, MSE, RMSE, and NSE to provide a comprehensive performance evaluation, focusing on both accuracy and error distribution. In the study [7], highlighted Precision and Accuracy for model reliability, while [6] emphasized R^2 and F1-Score, focusing on balancing precision and recall. The authors of [8] and [18] stressed the importance of MAE and RMSE to minimize forecasting errors. Meanwhile, [19] and [20] included NSE and F1-Score to evaluate how well models replicate observed data while balancing precision and recall. Overall, these studies use a wide range of metrics, offering a holistic understanding of model robustness and performance, enhancing climate prediction methodologies. Table 2 shows the performance metrics used to assess the models.

Table 1. Machine Learning and Deep Learning Techniques Used

References	SVM/ SVR	RF	LR	Lasso regression	NB	KNN	Decision Tree	BRT	K-Mean
[6]	/	/			/		/		
[7]			/					/	
[8]		/							/
[9]			/						
[17]	/	/							
[18]	/		/	/					
[19]	/		/						
[21]	/		/	/	/	/	/		
[22]	/	/						/	

Table 2. Comparison of Performance Metrics Used

References	Precision	R^2	F1-Score	Accuracy	MAE	MSE	RMSE	NSE
[6]	/		/					
[7]							/	
[8]			/					
[9]		/					/	
[17]				/		/	/	/
[18]		/					/	
[19]							/	
[20]	/		/	/				

3. RESEARCH METHODOLOGY

The research methodology section describes the steps undertaken to process the climate changes through the historical data of Malaysia by utilizing ML. The data is obtained from Visual Crossing Weather Data, preprocessed and features are extracted from the data as well. Then, three ML models are constructed: SVR, RFR, and LR. The models are then evaluated for performance using MAE, MSE and RMSE. Figure 1 shows the overview of the research methodology.

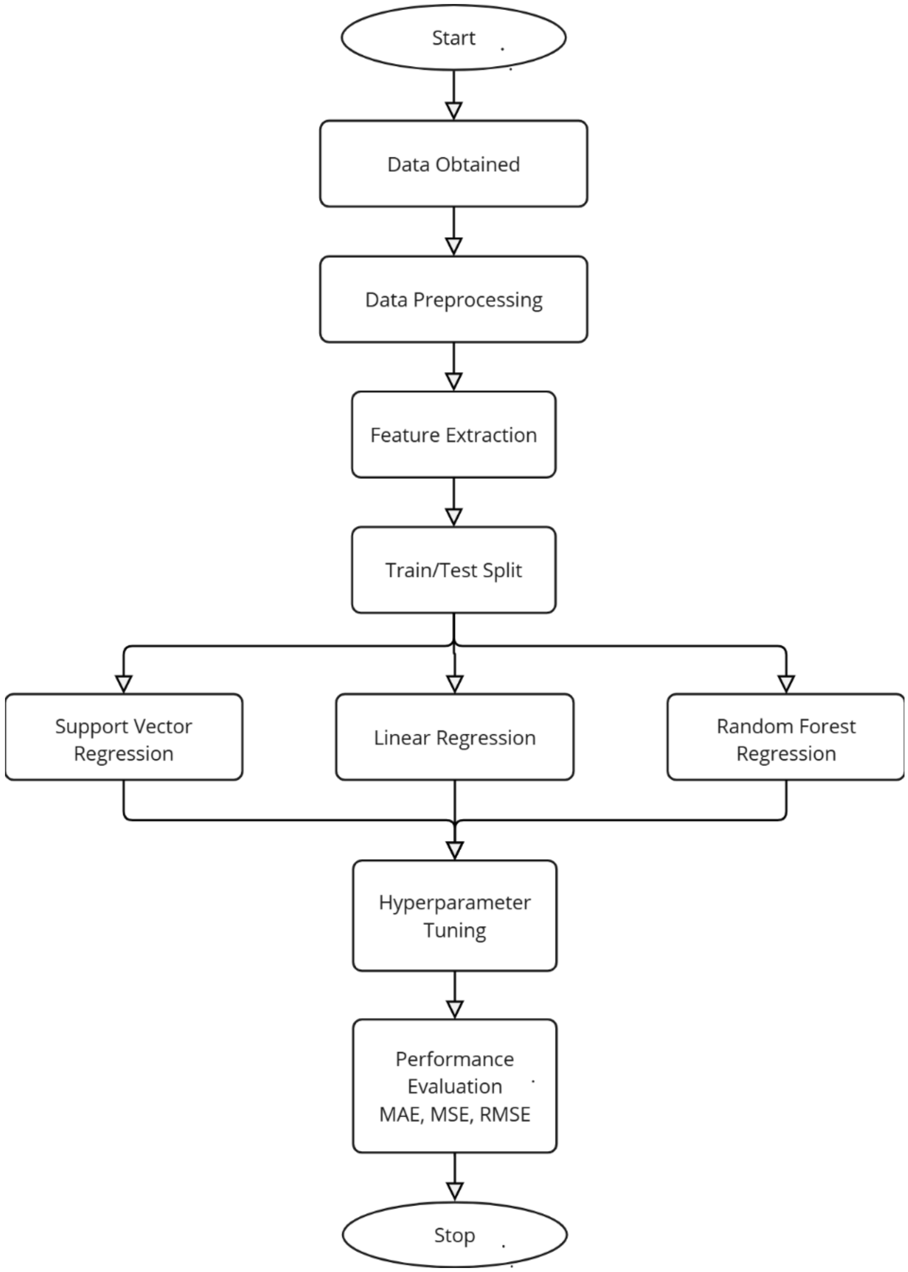

Figure 1. Research Methodology

3.1 Data Obtained

The Data is obtained from the website Visual Crossing Weather Data [23]. The Visual Crossing Weather dataset provides extensive weather and precipitation data, including historical weather summaries and forecasts. Key elements include maximum, minimum, and mean temperatures, precipitation amounts, wind speed and direction, humidity levels, and atmospheric pressure. The dataset is structured in a consistent tabular format allowing for easy access and analysis. Data connected to Air Surface Temperature and Precipitation have been collected. For this data collection, it has been collected yearly based one by one from the year 2014 to 2024. As it is limited to download the dataset from the website directly. Hence the data have been downloaded separately then merged into one dataset. Table 3 shows the dateset description.

Table 3. Dataset Description

Element	Description
name	Country
datetime	Date and Time
tempmax	Maximum Temperature
tempmin	Minimum Temperature
temp	Temperature (or mean temperature)
dew	Dew Point
feelslike	Feels like
precip	Precipitation
precipprob	Precipitation chance
precipcover	Precipitation cover
preciptype	Precipitation type
windspeed	Wind speed
windgust	Wind gust
winddir	Wind direction
visibility	Visibility
cloudcover	Cloud cover
humidity	Relative humidity
pressure	Sea level pressure
solarradiation	Solar radiation
solarenergy	Solar energy
uvindex	UV index
severerisk	Severe Risk
sunrise	Sunrise time
sunset	Sunset time
moonphase	Moonphase
icon	A weather icon
conditions	Short text about the weather
description	Description of the weather for the day
stations	List of weather stations sources

3.2 Data Preprocessing

Preprocessing or preparing the data: Preprocessing is one vital stage in a machine-learning pipeline that deals with cleaning and refining raw data so it is ready for analysis. This step involves fixing mistakes and removing unwanted noise; this will boost accuracy and reliability in the dataset by a great margin. This helps in excluding the irrelevant information and noise for the effective extraction of features; thus, it enhances the effectiveness of models. A well-structured dataset allows data engineers and researchers to apply machine-learning algorithms more effectively, which can improve the discovery of patterns and the forecasting of future events [6], [7].

The first task of data preprocessing is its cleaning, which encompasses the deletion of missing values, removal of duplicate observations, elimination of unwanted outliers and assortment of structural errors. In this case, the missing values can be handled by either deleting the columns of the null entries or putting some estimated values based on other observations. Next, data transformation: this process translates features into a more usable format rather for the analysis phase. This can be decomposing categorical variables into a numerical type or using parts of a datetime data source during the construction of other features. Sklearn developed a number of important methodological approaches to these activities, including the dataset division and the train_test_split, of course, StandardScaler and MinMaxScaler for normalization. The use of such approaches in their combined form allows the entire dataset to be well structured and well cleaned up in anticipation of the ML application models that follow.

3.3 Feature Extraction

Feature extraction in ML helps to optimally develop the model since raw data is converted into features that are more significant. Feature extraction mechanism promotes a reduction of overfitting that is by taking attention from the model on possible noisy data and so boosting the overall accuracy. Generally, feature extraction makes it possible for the models to quickly determine important patterns which yield efficient and accurate predictions. In order to enhance the model performance, dimensionality reduction via Principal Component Analysis (PCA) was performed using sklearn.decomposition module, which succeeded in capturing a large part of the data variability. According to [9], PCA does this by decreasing the dimensionality of the data and replacing the original features with fewer linear combinations of them that are less in number. Including only the most impactful variables in the model will guarantee accurate conclusions devoid of bias, hence improving the ML models.

3.4 Model Building

In the model building phase of the research, we carry out a number of critical steps which include dataset splitting, applying different models such as SVR, RFR, and LR. Once the data has been gathered, the first step is to split the preprocessed and feature extracted data into training and test dataset. The percentage of the training data is 80%, while the remainder for testing the data. With this split, the models can focus on learning a great chunk of data with some set aside to test and give unbiased perspectives of the expected outcomes of the developed models [6].

Next step is the training of the models, after dataset split. SVR is of great relevance because, by mapping the input features into a higher dimensional space, it can learn much more complex non-linear relationships. In this regard, SVR is able to identify the hyperplane with the smallest prediction error which enhances the accuracy of the model. On the other hand, employing an ensemble approach, which builds multiple decision trees during training and then averages their predictions, RFR works. This is beneficial as it improves robustness and lowers the chances of overfitting hence can work well on complicated data containing lots of features [8]. Despite its basic nature, LR can equally be taken as a good reference model because of its simplicity and low resource usage in building linear relationships.

These regression techniques have been chosen due to their respective capabilities in dealing with the climate data. It is able to capture complicated relationships as SVR utilizes the kernel trick which enables it to be efficient in high dimensional spaces. Due to the use of an ensemble learning approach in Random Forest, the problems of overfitting that are associated with a single decision tree are easily eliminated thus making it applicable in a dataset with high variability. A simpler model for LR is essential in this case as it provides a benchmark for the other complex models and allows for proper assessment of model usefulness.

3.5 Hyperparameter Tuning

Hyperparameters have also been known to have an impact on the factors that affect a model's performance which in this case is quite important. For each model, techniques such as Grid Search or Random Search are often utilized in full to ascertain as many hyperparameter combinations that would be suitable. Using this process, the most optimal settings that improve predictability without overfitting the model are established [7]. By performing hyperparameter optimization, we are able to improve model performance when the training data is different from that of the test data significantly.

3.6 Performance Evaluation

After the models have been built, it is important to assess their accuracy using various assessment techniques. The first type of evaluation used in this work is the MAE, MSE, and the RMSE, among others. The MAE measures average errors in prediction in a simple way. In MSE, when predicting and measured values are squared, squared loss focuses greater weight on bigger errors. RMSE facilitates computing the errors of various models on a common scale and depicting the errors more clearly. With the help of these performance metrics, it is possible to evaluate the ability of any models to predict temperature and precipitation in Malaysia [9].

The MAE is one of the most popular metrics in regression analysis where it quantifies the averageness of the errors in a particular set of forecasts or predictions without regard for their sign. MAE's formula can be expressed as the total of the absolute differences of actual values from the predicted values over N where N is the total number of observations. This is expressed in Equation (1) as

$$MAE = (1/n)\ \Sigma(i=1 \text{ to } n)\ |y\_i - \hat{y}\_i| \tag{1}$$

where n denotes the number of observations, yi is the real value of i-th observation Yi. Instead of cancelling positive and negative errors together as in the case of MSE, Absolute errors are preferred since it gives a better picture of average error. MAE is straightforward and makes sense conceptually because it is expressed in the same unit as the dependent variable. Hence MAE provides a simple way of modelling predictive accuracy.

MSE is one of the accepted model evaluation measures related to regression models, and this is used to find the average of the squared difference between the actual values and predicted ones. Moreover, the Equation (2) also demonstrates how to perform MSE and the following:

$$MSE = \Sigma(yi - pi)2n \tag{2}$$

In this case, yi is the true value; pi is the estimated value of the ith observation and n is the ith observation. Conclusively, MSE is a criterion which indicates the ability of a model to predict observation values of the target variable with some randomness. It assesses the accuracy of the model through MSE which is closely related to the better model performance and is indicated by less MSE value. Milder errors are on average squared less, thus more pronounced ones receive greater relative emphasis. Consequently, it is effective in differentiating between models prone to large errors or outlier values. Despite many benefits, there are also costs associated with the measure; in this case, first, the outcome is always in squared units which give an unusual interpretation compared to MAE.

The RMSE formula, expressed as in Equation (3). Here, Pi represents the predicted values, Oi denotes the actual observed values, (Pi−Oi) and calculates the residuals or errors. Squaring these differences eliminates negative values and emphasizes larger discrepancies, while Σ denotes the summation of all squared errors across observations. Dividing by n (the total number of observations) yields the MSE, and taking the square root converts it back to the original unit of measurement. RMSE provides a clear measure of how well the model performs, with lower values indicating better predictive accuracy.

$$RMSE = \text{sqrt}\ [(\Sigma(Pi - Oi)^2) / n] \tag{3}$$

4. RESULTS AND DISCUSSIONS

Figure 2 shows a comparison of performance metrics from different ML models used to predict temperature trends in Malaysia. The predictions made by the untuned models regarding the temperature demonstrate major inconsistencies in performance for several of the algorithms. In particular, SVR model encountered a MAE of 0.2789 and therefore exhibited relatively high prediction errors in contrast to other models. The LR model was stronger than RFR and SVR with the MAE of 0.0883 which allowed it to be very useful in modeling the trends in temperature data. On the other hand, the RFR model had an MAE of 0.3514. This means that RFR is best suited for predicting an entirely different task. It was evident from the Figure 2 that each model has its strengths but in conclusion the Reliability of LR model was the most dependable for the estimating temperature trend for Malaysia region. This encourages the use of LR in this research area since it shows consistent results regardless of the parameters. On the other hand, it seems that SVR and RFR, because of their underperformance, require more tuning and optimization for improved prediction performance.

Figure 3 shows a comparison of performance metrics for untuned ML models used to predict precipitation trends. The performance of the untuned precipitation prediction models reveals that there is a wider error throughout the algorithms evaluated. One of the untuned models with the SVR algorithm had a MAE of 5.0407, the best performance to have been achieved so far by the untuned models but still reveals significant prediction errors to be had. The MAE in an index of accuracy representing the average absolute difference between forecast and observed values was 6.

2577 for LR model and for RFR it was 6.3726 which shows their prediction is not as precise for the level of precipitation. These results suggest that there is none of the untuned models for the specific purpose of precipitation forecasting that has a high level of success, with SVR performing slightly better in relation to the other models. The results indicate that precipitation forecasting is unsatisfactory in the case of untuned models. This analysis in turn, highlights the importance of further improvements and tunings on these models to increase their forecast accuracy of precipitation trends to a decent level.

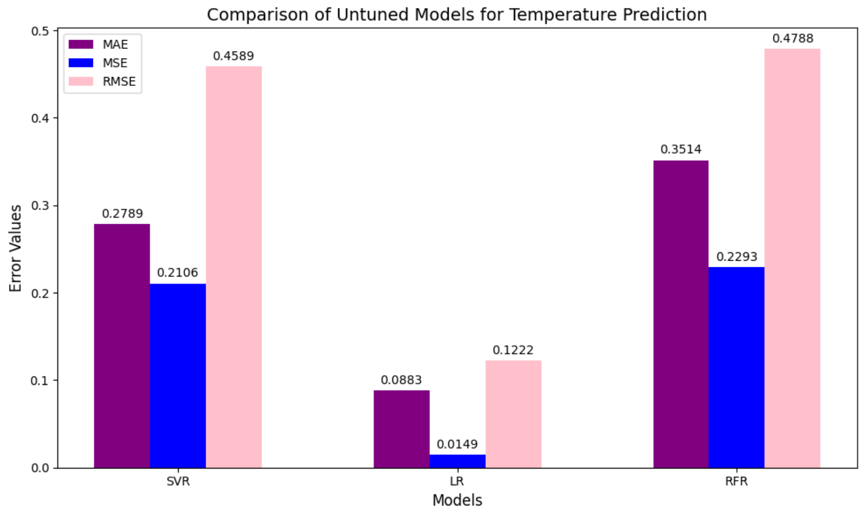

Figure 2. Comparison of Untuned Models for Temperature Prediction

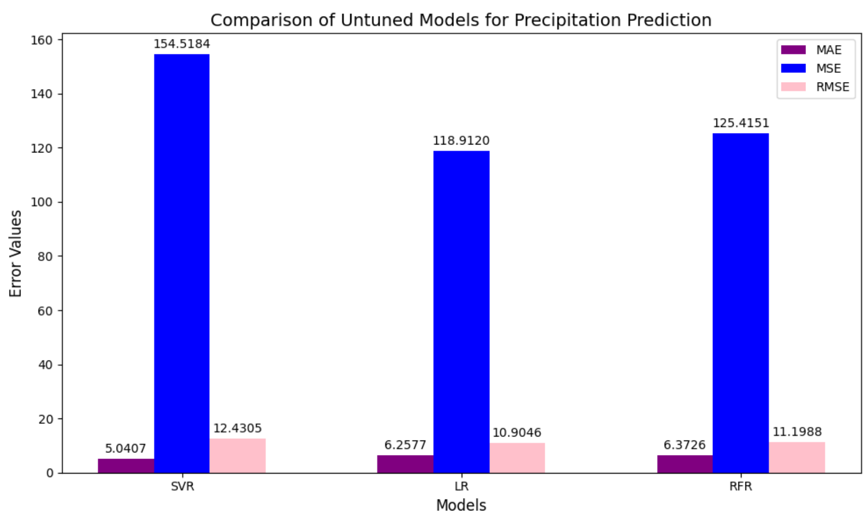

Figure 3. Comparison of Untuned Models for Precipitation Prediction

Figure 4 shows the results after hyperparameter tuning. The performance of all algorithms increased noticeably after the models were tuned for temperature prediction. The MAE for the SVR model that was tuned was found to be 0.2442, which is a significant improvement from the untuned SVR model. The LR model, on the other hand, was still the best one since it was able to post an MAE of 0.0894 and retains its strong ability to forecast temperatures even after tuning. It is also observed that there was no change in the performance of the RFR model which remained at 0.3514 after the tuning; this shows that the tuning didn't improve its ability to predict in this case. It can be seen from the results that while SVR and LR models were enhanced with hyperparameter tuning, RFR model did not improve with tuning. In all these cases, the tuned models, particularly the LR model, show a significant improvement in the performance of the predictor. The strong predictive results of the LR model help emphasize their reliability as temperature trends prediction.

Figure 5 shows a comparison of performance metrics for tuned ML models used to predict precipitation trends. Unlike in the case of temperature forecasts, the tuned models for the prediction of precipitation had different levels of improvement over the baseline models. But there was still low accuracy in the adjusted models. The tuned SVR model recorded a MAE of 4.9603, which is an improvement over the performance of the untuned SVR model only, but still, error margins are considerably large. The tuned performance of the LR model showed an MAE of 6.2206, slightly higher than that of the untuned version, and reveals an inability to predict the precipitation data reliably. RFR's tuned performance gave an MAE of 6.5624, which is a decrease in accuracy as compared to the untuned state. It is evident from these results that while tuning may improve some models' performances, it does not imply that all algorithms will show significant improvements.

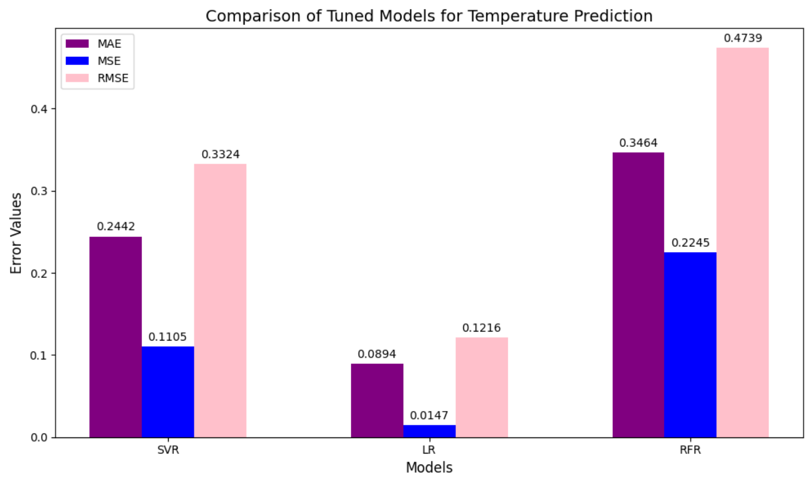

Figure 4. Comparison of Tuned Models for Temperature Prediction

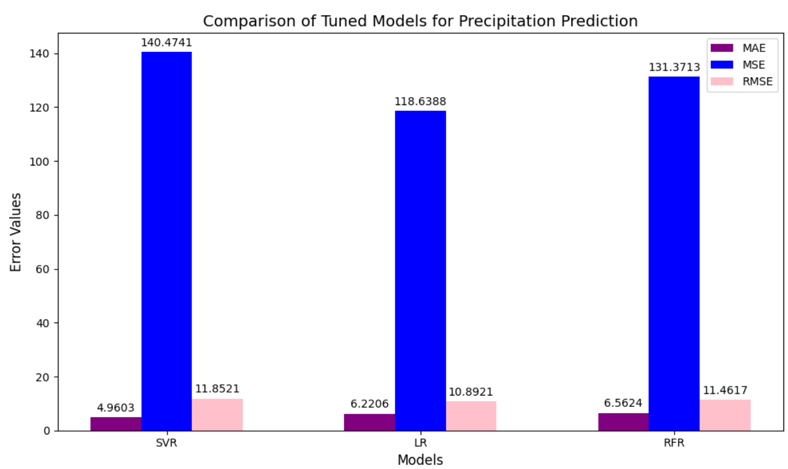

Figure 5. Comparison of Tuned Models for Precipitation Prediction

In Table 4, the performance evaluation metrics of several ml models in predicting the temperature after hyperparameter tuning and the ones which were not tuned are compared. After going through the tuning phase, one can even say that the model performance went through some positive changes as well as adverse changes. The on the other hand, improved from 0.2789 to 0.2442 for SVR model meaning that the regression model was much more accurate than before. On the contrary, the MAE of the LR model rose from 0.0883 to 0.0894 meaning that not much was achieved through tuning. In contrast, the So-called RFR model had a Mean Absolute Error of 0.3514 but even after tuning the algorithm it suggested that hyperparameter tuning did not help this algorithm in any way. These results suggest that there is a need to keep tuning the models since the predictive accuracy improves however, not all algorithms respond to tuning in the same fashion.

Table 4. Performance Comparison of Untuned and Tuned Models for Temperature

Model	MAE		MSE		RMSE	
	Untuned	Tuned	Untuned	Tuned	Untuned	Tuned
SVR	0.2789	0.2442	0.2106	0.1105	0.4589	0.3324
LR	0.0883	0.0894	0.0149	0.0147	0.1222	0.1216
RFR	0.3514	0.3514	0.2293	0.2293	0.4788	0.4739

Table 5 shows the performance comparison of untuned and tuned models for precipitation. The comparison of performance measures among untuned and optimized models for precipitation forecasting shows mixed outcomes pertaining to the enhancement of predicting accuracy. After optimization, a SVR model managed to reduce the MAE value from 5.0407 to 4.9603. For the ajuste degree, this is quite modest in the context of the fully absolute error level of the forecasts of precipitation. As a result of adjustments entering techniques, the pre-tuning MAE value of the LR model of 6.2577 was decreased to 6.2206, as a result of all these sequences, the MAE values were rather high in predictions. Adjustments for RFR performance were not as successful, the MAE in the training set increased to 6.5624 from 6.3726. These findings allow the conclusion that tuning is beneficial and applicable to certain configurations and datasets but, as is often the case, does not work 'homogeneously'.

Table 5. Performance Comparison of Tuned and Untuned Models for Precipitation

Model	MAE		MSE		RMSE	
	Untuned	Tuned	Untuned	Tuned	Untuned	Tuned
SVR	5.0407	4.9603	154.5184	140.4741	12.4305	11.8521
LR	6.2577	6.2206	118.9120	118.6388	10.9046	10.8921
RFR	6.3726	6.5624	125.4151	131.3713	11.1988	11.4617

These results suggest that there is a need to keep tuning the models since the predictive accuracy improves however, not all algorithms respond to tuning in the same fashion. This is consistent with recent works by Goh et al. [10], which states that tailored predictive models can play an important role in solving specific climate change issues and Ong et al. [11] show that accurate weather forecasting can help improve health management systems with the use of machine learning. All of these facts together give a rationale to the need for further research into optimization techniques to improve model performance.

The evaluation of the predictive models for temperature forecasting leads us to the concluding remark that the LR model has surpassed all the other models, given that in its tuned version it recorded the lowest MAE of 0.0883. Such accuracy achieved on a consistent basis, places LR as a very effective temperature forecasting model. SVR did improve post-resolution tuning. However, LR all the while was still reigning as superior. As the contrasting case, precipitation forecasts were not that favorable in all models including LR where SVR has remained dominant over it and RFR with respect to MAE. As such, the findings point out that LR remains the most appropriate method with respect to models to use while performing the temperature forecasting due to its fair predictive skills.

5. CONCLUSION

This study demonstrates the significant advantages of employing ML techniques for climate change predictions in Malaysia, revealing that LR outperforms traditional statistical methods in forecasting temperature and precipitation trends. The study employed Support Vector Regression Analysis, Random Forest Regression, and LR in analyzing the previous climate data for forecasting future temperature and precipitation trends. Results portrayed that LR model provided the best outcomes in dealing with forecasting reliability as it captures more complicated data relationships as evidenced by performance metrics such as MAE, MSE and RMSE.

In summary, the results of this research advocate for the integration of advanced ML methodologies into climate adaptation strategies. By leveraging these predictive capabilities, stakeholders can better anticipate and mitigate climate-related risks, thereby enhancing Malaysia's resilience to climate change. This study not only contributes to the existing body of knowledge but also emphasizes the importance of adopting innovative analytical tools to inform effective decision-making processes in addressing the pressing challenges posed by climate change.

ACKNOWLEDGEMENT

The authors would like to thank the two anonymous reviewers who have provided valuable suggestions to improve the article.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Anishalache Subramanian: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation;
Naveen Palanichamy: Project Administration, Supervision, Writing – Review & Editing;
Ng Kok Why: Project Administration, Writing – Review & Editing;
Sandhya Aneja: Review & Editing

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

REFERENCES

[1] M. F. Abdullah, M. Z. Mat Amin, M. F. Mohamad, M. Mohamad Ideris, Z. Zainol, and N. Y. Yussof, "N-HyDAA - Big Data Analytics for Malaysia Climate Change Knowledge Management," pp. 10–1. doi: 10.29007/5b3v.

[2] A. H. Ali and R. Thakkar, "Climate Changes through Data Science: Understanding and Mitigating Environmental Crisis," *Mesopotamian Journal of Big Data*, vol. 2023, pp. 125–137, Dec. 2023, doi: 10.58496/MJBD/2023/017.

[3] L. Y. Leng, O. H. Ahmed, and M. B. Jalloh, "Brief review on climate change and tropical peatlands," *Geoscience Frontiers*, vol. 10, no. 2, pp. 373–380, Mar. 2019, doi: 10.1016/j.gsf.2017.12.018.

[4] S. Das, C. Chatterjee, and S. Chakraborty, "A Machine Learning Approach to Re-Classification of Climate Zones Based on Multiple Rain Features Over India," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Jul. 2019, pp. 7752–7754. doi: 10.1109/IGARSS.2019.8898422.

[5] S. Vázquez-Ramírez, M. Torres-Ruiz, R. Quintero, K. T. Chui, and C. Guzmán Sánchez-Mejorada, "An Analysis of Climate Change Based on Machine Learning and an Endoreversible Model," *Mathematics*, vol. 11, no. 14, p. 3060, Jul. 2023, doi: 10.3390/math11143060.

[6] S. Zainudin, D. S. Jasim, and A. Abu Bakar, "Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction," *Int J Adv Sci Eng Inf Technol*, vol. 6, no. 6, p. 1148, Dec. 2016, doi: 10.18517/ijaseit.6.6.1487.

[7] B. S. Sidhu, Z. Mehrabi, N. Ramankutty, and M. Kandlikar, "How can machine learning help in understanding the impact of climate change on crop yields?," *Environmental Research Letters*, vol. 18, no. 2, p. 024008, Feb. 2023, doi: 10.1088/1748-9326/acb164.

[8] T. Ferdousi *et al.*, "A Machine Learning Framework to Explain Complex Geospatial Simulations: A Climate Change Case Study," in *2023 Winter Simulation Conference (WSC)*, IEEE, Dec. 2023, pp. 2990–3001. doi: 10.1109/WSC60868.2023.10408406.

[9] R. Wamanse and T. Patil, "Analysis of various climate change parameters in India using machine learning," Jan. 2022.

[10] J.-L. Goh, S.-B. Ho, and C.-H. Tan, "Weather-Based Arthritis Tracking: A Mobile Mechanism for Preventive Strategies," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 210–225, Feb. 2024, doi: 10.33093/jiwe.2024.3.1.14.

[11] W.-X. Ong, S.-B. Ho, and C.-H. Tan, "Enhancing Migraine Management System through Weather Forecasting for a Better Daily Life," *Journal of Informatics and Web Engineering*, vol. 2, no. 2, pp. 201–217, Sep. 2023, doi: 10.33093/jiwe.2023.2.2.15.

[12] H. L., "Tackling Climate Change with Machine Learning," 2021.

[13] R. A. A. Rashid, P. N. E. Nohuddin, and Z. Zainol, "Analyzing Climate Variability in Malaysia Using Association Rule Mining," *International Journal of Engineering & Technology*, vol. 7, no. 4.34, p. 394, Dec. 2018, doi: 10.14419/ijet.v7i4.34.26881.

[14] H. A. Rahman, "CLIMATE CHANGE SCENARIOS IN MALAYSIA: ENGAGING THE PUBLIC," *International Journal of Malay-Nusantara Studies*, vol. 1, no. 2, 2018.

[15] N. A. Ismail and W. Z. Wan Zain, "Statistical Analysis of Rainfall Data in Peninsular Malaysia using Strong Transformation of Cubic Spline Analysis," *Jurnal Kejuruteraan*, vol. si5, no. 2, pp. 237–240, Nov. 2022, doi: 10.17576/jkukm-2022-si5(2)-26.

[16] M. N. I. Kassim *et al.*, "Climate Change Research in Malaysia: A Scientometric Analysis," *Journal of Scientometric Research*, vol. 12, no. 3, pp. 534–543, Nov. 2023, doi: 10.5530/jscires.12.3.051.

[17] N. A. F. Sulaiman, S. M. Shaharudin, S. Ismail, N. H. Zainuddin, M. L. Tan, and Y. Abd Jalil, "Predictive Modelling of Statistical Downscaling Based on Hybrid Machine Learning Model for Daily Rainfall in East-Coast Peninsular Malaysia," *Symmetry (Basel)*, vol. 14, no. 5, p. 927, May 2022, doi: 10.3390/sym14050927.

[18] V. H., "Climate Change Analysis Using Machine Learning," *International Journal of Science and Research (IJSR)*, vol. 9, no. 8, pp. 973–977, 2020.

[19] K. Chattrairat, W. Wongseree, and A. Leelasantitham, "Comparisons of Machine Learning Methods of Statistical Downscaling Method: Case Studies of Daily Climate Anomalies in Thailand," *Journal of Web Engineering*, Jul. 2021, doi: 10.13052/jwe1540-9589.2057.

[20] V. Shevchenko *et al.*, "Climate Change Impact on Agricultural Land Suitability: An Interpretable Machine Learning-Based Eurasia Case Study," *IEEE Access*, vol. 12, pp. 15748–15763, 2024, doi: 10.1109/ACCESS.2024.3358865.

[21] P. Mangal, A. Rajesh, and R. Misra, "Big Data in Climate Change Research: opportunities and Challenges," in *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, IEEE, Jun. 2020, pp. 321–326. doi: 10.1109/ICIEM48762.2020.9160174.

[22] T. Ladi, S. Jabalameli, and A. Sharifi, "Applications of machine learning and deep learning methods for climate change mitigation and adaptation," *Environ Plan B Urban Anal City Sci*, vol. 49, no. 4, pp. 1314–1330, May 2022, doi: 10.1177/23998083221085281.

[23] Visual Crossing, "Weather Data Documentation." Accessed: Sep. 23, 2024. [Online]. Available: https://www.visualcrossing.com/resources/documentation/weather-data/weather-data-documentation/

BIOGRAPHIES OF AUTHORS

	Anishalache Subramanian is a final-year undergraduate student pursuing a Bachelor of Computer Science (Hons) in Data Science at Multimedia University, Malaysia. She has demonstrated strong problem-solving and leadership skills through her academic and extracurricular activities. She aspires to build a career in data science or machine learning, contributing to innovative solutions for real-world challenges. She can be contacted at anishalache@gmail.com.
	Palanichamy Naveen is a Senior Lecturer at the Faculty of Computing and Informatics, Multimedia University. She joined the faculty after completing her PhD at Curtin University, Malaysia. She earned her Bachelor of Engineering (CSE) and Master of Engineering (CSE) from Anna University, India. Her research interests include smart grids, cloud computing, machine learning, deep learning, generative AI and recommender systems. She is actively involved in several research projects funded by Multimedia University. She can be contacted at p.naveen@mmu.edu.my.
	Kok-Why Ng is an Associate Professor in the Faculty of Computing and Informatics (FCI) in Multimedia University (MMU), Malaysia. He did his B.Sc. (Math) in USM, Penang, and his M.Sc (IT) and Ph.D (IT) in MMU, Malaysia. His research interests are in Recommender System, 3D Geometric Modeling and Animation. He is also active in some research projects related to artificial intelligence, deep learning and human blood cells. He can be contacted at kwng@mmu.edu.my.
	Sandhya Aneja is an Assistant Professor of Computer Science at Marist University in Poughkeepsie, New York, since Sept 2022. Previously, she worked at the University of Brunei and Delhi University. She has 20 years of teaching experience. She has published over 44 peer-reviewed publications and one US patent focusing on computer networks, security, artificial intelligence, and deep learning. She has worked in various countries, including India, Brunei, and the United States, where she has mentored students from diverse backgrounds to enhance accessibility in education. Her contributions over the years reflect her commitment to teaching, research, service, and mentorship. She can be contacted at sandhya.aneja@marist.edu.