Journal of Informatics and Web Engineering

Vol. 4 No. 1 (February 2025) eISSN: 2821-370X

Predicting Short-Range Weather in Tropical Regions Using Random Forest Classifier

Sellappan Palaniappan[1]*, Rajasvaran Logeswaran[2], Anitha Velayutham[3], Bui Ngoc Dung[4]

[1]Corporate, HELP University, 5, Jalan Sri Semantan 1, Off, Jalan Semantan, Bukit Damansara, 50490 Kuala Lumpur, Malaysia.
[2,3]Faculty of Computing and Digital Technology, HELP University, 5, Jalan Sri Semantan 1, Off, Jalan Semantan, Bukit Damansara, 50490 Kuala Lumpur, Malaysia.
[4]University of Transport and Communications, No.3 Cau Giay Street, Lang Thuong Ward, Dong Da District, Hanoi, Vietnam.
**corresponding author: (sellappan.p@help.edu.my; ORCiD: 0009-0009-1168-2864)*

Abstract - In this paper, we present a Random Forest classifier machine learning model for predicting short-range weather in in tropical regions like Malaysia. Our model uses environmental factors such as temperature, humidity, wind speed, and cloud cover to predict weather conditions like clear skies, rain, and thunderstorms. Tropical weather, influenced by high humidity, fluctuating temperatures, and frequent rainfall, present unique challenges for forecasting accurately. To address these challenges, we trained a Random Forest classifier on a synthetic (simulated) dataset comprising 1,500 samples, each representing a specific weather scenario. Our model achieved an accuracy of 98.66% in predicting short-term weather conditions, identifying cloud cover, precipitation intensity, and humidity as the most influential factors. Our model's high accuracy demonstrates its potential for predicting short-range weather conditions in tropical regions. Potential applications of the model spans sectors like agriculture, energy, tourism, disaster management, and public health. In agriculture, the model can be used to optimize irrigation schedules and crop management. In the energy sector, it can be used to optimize energy production and distribution. In disaster management, it can alert residents of impending bad weather, so they are more prepared. In the health sector, it can provide timely weather alerts and assist those who are more prone to arthritis and migraine attacks. We can enhance the model by using real-world data and regional customization.

Keywords— Random Forest, Machine Learning Model, Short-Range Weather Prediction, Tropical Regions, Disaster Management

Received: 02 September 2024; Accepted: 13 November 2024; Published:16 February 2025

1. INTRODUCTION

Predicting short-range weather accurately is crucial for many sectors such agriculture, energy, tourism, and public health [1]-[3]. Environmental factors like temperature, wind speed, humidity, precipitation, and cloud cover influence if a region experiences sunny, cloudy, windy, or rainy weather[4],[5]. These factors are often unpredictable, especially

MULTIMEDIA UNIVERSITY

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2025.4.1.2

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

in tropical regions like Malaysia [6], thus that makes accurate short-range weather prediction more challenging.

Using traditional methods to forecast short-range weather in tropical regions, though useful, has not very effective [7],[8]. We can now use advanced machine learning models to predict short-range weather more accurately [9]. Machine learning can model and capture more complex interactions of environmental factors [10],[11]. Our study uses a Random Forest classifier machine learning model to predict short-range weather in tropical regions using synthetic (simulated) environmental factors like temperature, humidity, precipitation, wind speed, and cloud cover [12].

Our study aims to answer the following research questions: (i) What environmental factors influence short-range weather in tropical regions?, (ii) Can Random Forest classifier machine learning model be used to predict short-range weather in tropical regions?, (iii) How does Random Forest classifier compare with traditional approaches in terms of prediction accuracy?, and (iv) Can this model be applied to benefit different sectors?

The following research objectives attempts to answer the above research questions:

- Identify environmental factors that influence short-range weather in tropical regions.
- Build a Random Forest machine learning model to predict short-range weather in tropical regions.
- Compare the performance of Random Classifier with traditional approaches in terms of prediction accuracy.
- Recommend the application of this model in different sectors.

2. LITERATURE REVIEW

Machine learning is now used widely in many sectors such as healthcare, biotechnology, agriculture, education, finance, and manufacturing. It is also used in predicting weather in tropical regions. Machine learning can capture complex interactions between various environmental factors such as rainfall, wind speed, humidity and cloud cover. One study [13] employed deep learning LSTM recurrent neural networks to predict short-term precipitation more accurately than traditional methods. Similarly, another study [14] used deep learning models to predict rainfall in tropical regions more accurately than the traditional methods. These studies demonstrate the effectiveness of deep learning models in capturing complex interactions between environmental factors. Another study [15] demonstrated the effectiveness of ensemble models over traditional methods for forecasting 0–6-hour precipitation. They attributed this to the superiority of ensemble models over traditional models, which are not effective in complex interactions.

Some are more sensitive to weather conditions. For example, bad weather can spike the incidence of migraine and arthritis attacks. We can reduce these incidences of attacks by predicting short-term weather more accurately. For example, [16] developed a mobile app that provided weather alerts to help those prone to arthritis and migraine attacks. These weather alerts were based on short-range weather forecasts retrieved from the OpenWeather API. Similarly, [17] created a migraine management app using short-range weather forecasts to predict the likelihood of migraine attacks. This shows that accurate prediction of short-range weather in tropical regions can help improve health by reducing the incidence of certain weather-related ailments.

Today, we generate electricity using fossil fuel, solar, hydro and wind power. Conserving energy is important. We do not want to generate more power than we need, and vice versa. So, regulating the supply and demand for power is important.

Weather influences the demand and consumption of electricity. Accurate short-range weather prediction is very useful in sectors like agriculture, tourism, and energy. One study [18] stressed the importance weather forecasting for regulating renewable energy production using solar and wind power. Another study [19] stressed the need for accurate weather prediction for managing agricultural crops more optimally.

As highlighted by [20], we can now use advanced machine learning techniques to predict short-range weather more accurately. This study uses Random Forest classifier machine learning model to predict short-range more accurately compared to traditional approaches.

3. RESEARCH METHODOLOGY

3.1 Model Architecture

Our study uses a Random Forest classifier model with synthetic data to predict short-range weather in tropical regions. We selected this algorithm because it can handle complex relationships present in the interaction of the various environmental factors such as wind, humidity and rainfall. Besides, it is resistant to overfitting which can affect the model's generalizability to new data. Another reason is its ability to rank feature importance.

We implemented the Random Forrest classifier model using Python and libraries like scikit-learn library. We tuned the following parameters to 100 decision trees (n_estimators = 100) and a random state of 42 for reproducibility.

Figure 1 shows the model's architecture, consisting of several modules (components): input, data preprocessing, feature extraction, Random Forest model, model training, performance evaluation, prediction, and visualization. The arrows indicate the flow of data through the system modules.

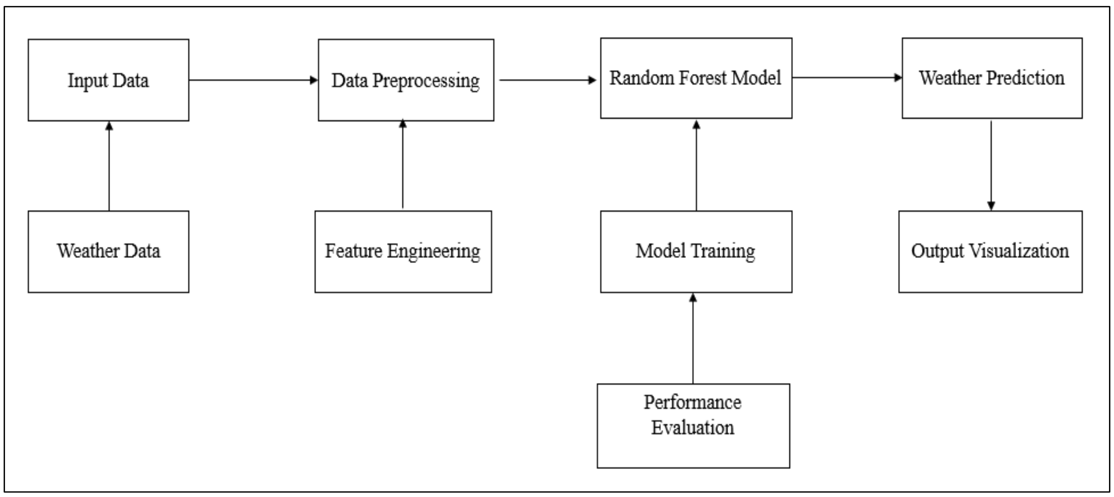

Figure 1. Model Architecture with Modules

3.2 Generation of Synthetic Dataset

Due to lack of availability of real-world data, we generated a realistic synthetic dataset of 1,500 samples to train and test our model. Each sample represents a specific weather scenario (not just a weather condition). We used the following environmental factors.

a) Temperature (°C)
b) Humidity (%)
c) Wind Speed (km/h)
d) Wind Direction (degrees)
e) Pressure (hPa)
f) Cloud Cover (%)
g) Precipitation Intensity (mm/h)
h) UV Index
i) Wind Gust (km/h)
j) Elevation (m)
k) Proximity to Coastline (km)
l) Sea Surface Temperature (°C)

We selected these factors because they are most relevant for short-range weather prediction in tropical regions [6]. The inclusion of environmental factors such as elevation and proximity to the coastline helps the model to account for local weather patterns.

Each sample simulates all the 12 factors for a specific weather scenario. The weather condition (target variable) for each sample falls into one of these six classes - Clear, Cloudy, Drizzle, Rain, Stormy, or Thunderstorm. These classes are typically used to simulate realistic tropical weather scenarios.

3.3 Data Preprocessing

We prepared the dataset using three steps: 1) We used feature scaling to normalize numerical data, which helps to balance the influence of each feature on the model's outcomes, 2) We used one-hot coding to convert non-numerical data, such as weather types, to a format suitable for machine learning, and 3) We split the dataset into 80% for training the model and 20% for testing the model as this is the common practice.

3.4 Model Implementation

We implemented a Random Forest classifier using Python's scikit-learn library. This is an ensemble technique that combines multiple decision trees to predict the most frequent class [9]. We used 100 decision trees (n_estimators = 100) and a fixed random state of 42 to ensure reproducibility.

The pseudocode for the implementation is shown in Algorithm 1. This algorithm takes in some environmental features such as temperature, humidity, wind speed and so on. The output of this algorithm is the predicted weather condition. Through the function GenerateSyntheticData, we randomly generate the dataset with the environment defined earlier (Line 1 to 6). Next, the the generated dataset will undergo the preprocessing stage (Line 7 to 11); typically splitting the data into 80% training and 20% testing. Then, it undergoes the TrainModel function based on Random Forest classifier (Line 12 to 15).

Algorithm 1: To Predict Weather Using Random Forest

```
Algorithm: Tropical weather prediction using Random Forest
Input: Environmental features (temperature, humidity, wind speed, etc.)
Output: Predicted weather condition

1. Function GenerateSyntheticData(n_samples):
        2. For each feature in feature_list:
                3. Generate random values within typical tropical ranges
        4. For each sample:
                5. Assign weather condition based on feature combinations
        6. Return dataset

7. Function PreprocessData(dataset):
        8. Normalize numerical features
        9. One-hot encode categorical variables
        10. Split data into training (80%) and testing (20%) sets
        11. Return X_train, X_test, y_train, y_test

12. Function TrainModel(X_train, y_train):
        13. Initialize RandomForestClassifier with n_estimators=100
        14. Fit model on training data
        15. Return trained_model

16. Function EvaluateModel(model, X_test, y_test):
        17. Make predictions on test data
        18. Calculate accuracy, precision, recall, F1-score
```

```
        19. Generate classification report
        20. Return evaluation_metrics

21. Function AnalyzeFeatureImportance(model, feature_names):
        22. Extract feature importances from model
        23. Sort features by importance
        24. Return sorted_feature_importances

25. Main:
        26. data = GenerateSyntheticData(1500)
        27. X_train, X_test, y_train, y_test = PreprocessData(data)
        28. model = TrainModel(X_train, y_train)
        29. evaluation_metrics = EvaluateModel(model, X_test, y_test)
        30. feature_importances = AnalyzeFeatureImportance(model, feature_names)
        31. Print evaluation_metrics and feature_importances
```

3.5 Model Evaluation and Analysis

We evaluated our model using the following metrics (see Algorithm 1 – Line 16 to 20):

- Accuracy – the overall percentage of correct predictions.
- Precision - the proportion of correct positive predictions within all predicted positives for each class.
- Recall - the proportion of correct positive predictions among all actual positive instances.
- F1-score - a balanced metric combining precision and recall.

To assess the influence of different environmental factors, we analyzed feature importance using the Random Forest's feature_importances_ function which ranks features based on their contribution to reducing impurity across the forest's trees, providing insights into which factors significantly impact tropical weather conditions (see Algorithm 1 – Line 21 to 24).

We tested the model's real-world applicability by creating five new distinct samples of weather scenarios. By applying the trained model to these new samples, we demonstrated its potential to generate short-term forecasts for specific weather scenarios in tropical regions.

4. RESULTS AND DISCUSSION

4.1 Model Performance

The Random Forest model demonstrated exceptional performance in predicting short-range weather conditions, achieving an overall accuracy rate of 98.66%. This high accuracy indicates the model's strong capability in distinguishing between different weather conditions in tropical regions.

Table 1 shows the detailed performance metrics for each weather condition. The classification report shows high precision and recall scores across most weather categories. The model achieved perfect precision and recall for Rain and Stormy conditions, which are critical for short-range forecasting in tropical regions. The lower precision for Clear conditions (0.43) coupled with perfect recall suggests that the model might be over-predicting this category, possibly due to the limited number of Clear samples in the dataset.

Figure 2 shows the confusion matrix of the model's weather conditions. The confusion matrix shows the model's strong performance, with most predictions concentrated along the diagonal, indicating correct classifications. The few misclassifications are primarily between similar weather conditions, such as Drizzle and Rain, which could be due subtle differences between these weather conditions.

Table 1. Model Performance for Each Weather Condition

Weather Condition	Precision	Recall	F1-score	Support
Clear	0.43	1.00	0.60	3
Cloudy	1.00	0.92	0.96	25
Drizzle	1.00	0.96	0.98	28
Rain	1.00	1.00	1.00	141
Stormy	1.00	1.00	1.00	53
Thunderstorm	1.00	0.98	0.99	50

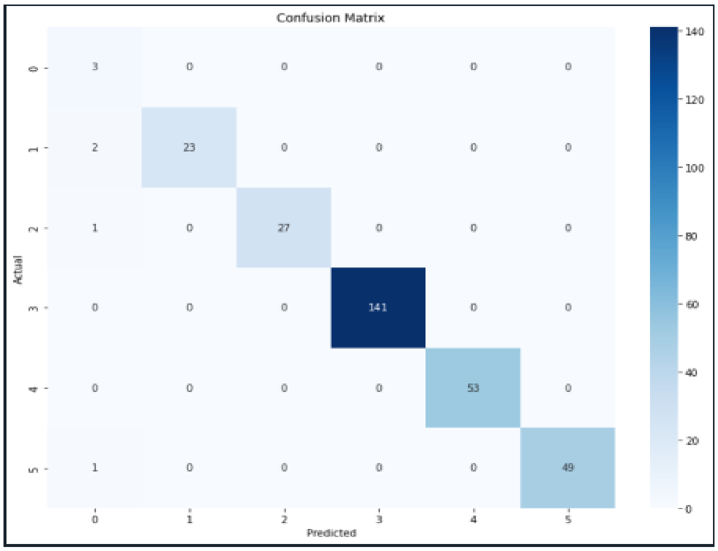

Figure 2. Confusion Matrix for The Model's Weather Conditions

4.2 Feature importance and feature correlation

Figure 3 shows the relative importance of each feature. The features that influence most in tropical weather are:

- Cloud cover
- Precipitation intensity
- Humidity
- Temperature
- Wind speed

This aligns with meteorological knowledge, where cloud cover and precipitation intensity are the primary factors in tropical weather, followed by humidity and temperature. The feature importance analysis provides some insights into the factors shaping weather conditions in tropical regions.

Figure 4 shows the feature correlation heatmap, which helps us understand the relationships between different environmental factors. In our case, the heatmap does not show high correlations between the environmental factors. The small positive correlations are between precipitation intensity and cloud cover, and between temperature and proximity to coastline. The small negative correlations are between wind direction and wind gust, and between temperature and precipitation intensity. These small correlations need further investigation.

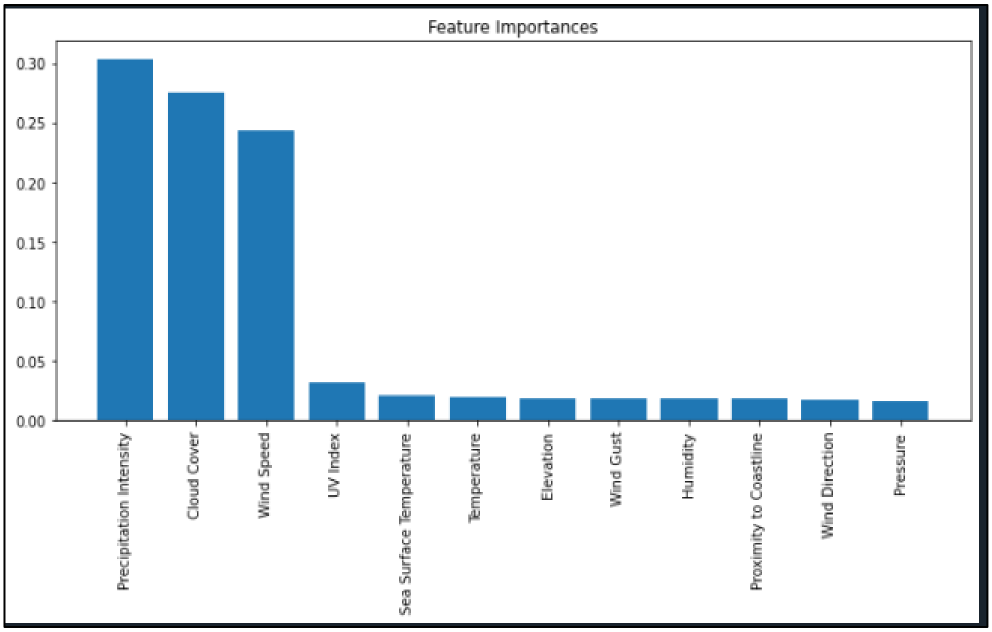

Figure 3. Relative Importance of The Features

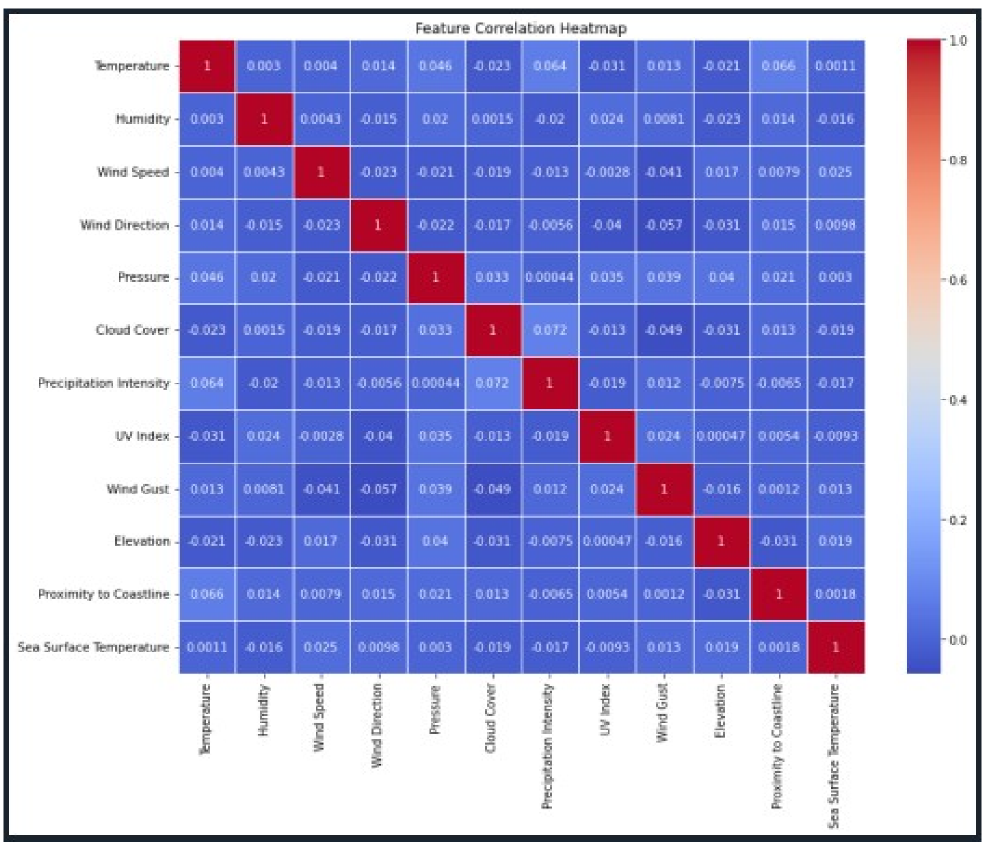

Figure 4. Feature Correlation Heatmap

4.3 Discussion

The high accuracy of our model shows that it can be used effectively to forecast short-range weather accurately. The model promises to benefit those sectors which are more subject to weather for their optimal functioning. For example, our model can help patients who are prone to weather-related attacks like arthritis by providing timely weather alerts. It can help farmers to optimize their crop harvests. It can help improve disaster preparedness, improve tourism planning, and improve energy conservation. Despite its potential, our model has several limitations: it is trained only on synthetic data, it is tailored only for tropical regions, it offers only short-range weather prediction, it has not considered other environmental factors.

5. CONCLUSION AND FUTURE WORK

This study demonstrated the potential of machine learning for improving short-range weather prediction in tropical regions like Malaysia. With an accuracy of 98.66%, the model promises potential benefits for weather-sensitive sectors like agriculture, energy, health and tourism.

Future work could use real-world data to train and test the model, provide localized weather prediction for smaller regions, and provide user-friendly apps for other sectors. These improvements will further enhance and broaden the model's applicability and make it more valuable.

ACKNOWLEDGEMENT

We thank the anonymous reviewers for the careful review of our manuscript.

FUNDING STATEMENT

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

AUTHOR CONTRIBUTIONS

Sellappan Palaniappan: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation, Final Manuscript Checking and Approval;
Rajasvaran Logeswaran: Project Administration, Writing – Review & Editing, Final Manuscript Checking and Approval;
Anitha Velayutham: Writing – Review & Editing, Final Manuscript Checking and Approval.
Bui Ngoc Dung: Writing – Review & Editing, Final Manuscript Checking and Approval.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

The paper follow The Committee of Publication Ethics (COPE) guideline.

No ethical issues. Synthetic data was used in the work.

REFERENCES

[1] T. Vovk, M. Kryza, and M. Werner, “Using random forest to improve EMEP4PL model estimates of daily PM2.5 in Poland,” *Atmos Environ*, vol. 332, p. 120615, Sep. 2024, doi: 10.1016/j.atmosenv.2024.120615.

[2] R. Pandey, M. Upadhya, and M. Singh, “Rainfall Prediction Using Logistic Regression and Random Forest Algorithm,” in *2024 IEEE International Conference on Computing, Power and Communication Technologies* (IC2PCT), IEEE, Feb. 2024, pp. 663–668. doi: 10.1109/IC2PCT60090.2024.10486681.

[3] M. Schnieder, “Using Explainable Artificial Intelligence (XAI) to Predict the Influence of Weather on the Thermal Soaring Capabilities of Sailplanes for Smart City Applications,” *Smart Cities*, vol. 7, no. 1, pp. 163–178, Jan. 2024, doi: 10.3390/smartcities7010007.

[4] S.-C. Hsu, A. K. Sharma, R. Tanone, and Y.-T. Ye, “Predicting Rainfall Using Random Forest and CatBoost Models,” Apr. 2024. doi: 10.11159/icgre24.146.

[5] S. Mathur, S. Kumar, and T. Choudhury, “Prognostication of Weather Patterns using Meteorological Data and ML Techniques,” *EAI Endorsed Transactions on Energy Web*, vol. 11, Apr. 2024, doi: 10.4108/ew.5648.

[6] C. L. Wong, R. Venneker, S. Uhlenbrook, A. B. M. Jamil, and Y. Zhou, “Variability of rainfall in Peninsular Malaysia”, *Hydrology and Earth System Sciences Discussions*, vol. 6, pp. 5471-5503, 2009, doi: 10.5194/hessd-6-5471-2009.

[7] M. T, S. S, Y. A, V. R. T, N. C, and G. P, “Hybrid Weather Forecasting: Integrating LSTM Neural Networks and Random Forest Models for Enhanced Accuracy,” *in 2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI*), IEEE, Apr. 2024, pp. 1–5. doi: 10.1109/RAEEUCCI61380.2024.10547997.

[8] V. Sharmila, A. R. Kannan, S. Kannadhasan, P. Sivakumar, and V. Vennila, Challenges in Information, Communication and Computing Technology. *CRC Press*, 2024.

[9] K.B. Jayanthi, K. Vanitha, K. Jeevanantham, S. Jayashree, A. Kaviya Mani, Assessing the efficacy of machine learning algorithms in weather forecasting: A comparative study of Erode weather data from local weather station. *CRC Press*, 2025.

[10] H. Chan and M. I. M. Wahab, “A machine learning framework for predicting weather impact on retail sales,” *Supply Chain Analytics*, vol. 5, p. 100058, Mar. 2024, doi: 10.1016/j.sca.2024.100058.

[11] Y. Shen, N. Liu, S. Zhang, X. Zhu, and N. An, “Research on the refinement of atmospheric weighted average temperature model in Xi’an based on machine learning,” *Advances in Space Research*, vol. 75, no. 1, pp. 353–366, Jan. 2025, doi: 10.1016/j.asr.2024.09.045.

[12] H. Chan and M. I. M. Wahab, “A machine learning framework for predicting weather impact on retail sales,” *Supply Chain Analytics*, vol. 5, p. 100058, Mar. 2024, doi: 10.1016/j.sca.2024.100058.

[13] A. A. Asanjan, T. Yang, K. Hsu, S. Sorooshian, J. Lin, and Q. Peng, “Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks”. *Journal of Geophysical Research: Atmospheres*, vol. 123, no. 22, pp. 12-543, 2018, doi: 10.1029/2018JD028375.

[14] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, “A review of wind power and wind speed forecasting methods with different time horizons”, In *Proceedings of the 2010 North American Power Symposium*, Arlington, 26-28 September 2010, pp. 1-8, doi: 10.1109/NAPS.2010.5619586.

[15] F. Sseguya , and K.-S. Jun, “Deep Learning Ensemble for Flood Probability Analysis”, *Water,* vol. 16, no. 21, p. 3092, 2024, doi: 10.3390/w16213092.

[16] 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 (JIWE)*, vol. 3, no. 1, pp. 210-225, 2024, doi: https://doi.org/10.33093/jiwe.2024.3.1.14.

[17] 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 (JIWE)*, vol. 2, no. 2, pp. 201-217, 2023, doi: https://doi.org/10.33093/jiwe.2023.2.2.15."

[18] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook", *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376-388, 2020, doi: 10.1109/OAJPE.2020.3029979.

[19] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review", *Sensors*, vol. 18, no. 8, p. 2674, 2018, doi: 10.3390/s18082674.

[20] L. Breiman, "Random forests", *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.

BIOGRAPHIES OF AUTHORS

	**Sellappan Palaniappan** is a Professor of Information Technology at HELP University, Malaysia. With over 30 years of academic experience, his current research interests are in the application of artificial intelligence and machine learning in diverse domains like cybersecurity, data analytics, healthcare, biotechnology, retail, education, agriculture, logistics, and sustainable development initiatives. He is also interested in quantum physics, DNA, neuroscience, and energy, frequency, and vibration for health and wholeness. He has published more than 100 scholarly research papers and authored several widely-used IT textbooks for college and university students. He may be contacted at sellappan.p@help.edu.my.
	**Rajasvaran Logeswaran, SMIEEE** is a Professor and Dean of Computing and Digital Technology at HELP University, Malaysia. With over 25 years of academic experience, his research interests are in medical image processing, artificial intelligence, data science and cybersecurity, with over 180 publications in books, peer-reviewed journals and international conference proceedings. He actively serves as a speaker at many international conferences, as well as volunteers as a judge in STEM and innovation competitions for schools at the local and international levels. He may be contacted at logeswaran.nr@help.edu.my.
	**Anitha Velayutham** is a Senior Lecturer at HELP University, Malaysia. With over 15 years of experience in IT education, she has cultivated a diverse teaching portfolio spanning cybersecurity, computing ethics, technopreneurship, and business development. As a dedicated mentor, she excels in guiding students through hands-on graduate projects, fostering their academic and professional growth. In addition to teaching, she oversees both internal and external graduate projects, ensuring students gain practical insights and real-world experience. She may be contacted at anithav@help.edu.my.

	**Bui Ngoc Dung** is a Senior Lecturer in Information Technology at the University of Transport and Communications, Vietnam. He holds a Ph.D. in Informatics from Malaysia University of Science and Technology and has completed postdoctoral research at TU Wien, Austria. His research focuses on machine learning, computer vision, biomedical applications, and structural health monitoring, with over 50 publications and numerous grants in applied AI for transportation. Dr. Dung also serves as a reviewer, technical committee member, and keynote speaker at international conferences.