Thunderstorm Prediction Model Using Hybrid Clustering and Machine Learning Approach

Shirley Anak Rufus
Institute of High Voltage and High Current (IVAT),
Faculty of Electrical Engineering,
University Teknologi Malaysia (UTM)
Johor Bahru, Malaysia
Department of Electrical and Electronics Engineering, Faculty of Engineering,
University Malaysia Sarawak (UNIMAS)
Kota Samarahan, Sarawak, Malaysia
rshirley@unimas.my

Noor Azlinda Ahmad
Institute of High Voltage and High Current (IVAT),
Faculty of Electrical Engineering,
University Teknologi Malaysia (UTM)
Johor Bahru, Malaysia
noorazlinda@utm.my

Zulkurnain Abdul-Malek
Institute of High Voltage and High Current (IVAT),
Faculty of Electrical Engineering,
University Teknologi Malaysia (UTM)
Johor Bahru, Malaysia
zulkurnain@utm.my

Noradlina Abdullah
Lightning & Earthing Unit,
TNB Research Sdn. Bhd.
Selangor, Malaysia
noradlina.abdullah@tnb.com.my

Nurul ‘Izzati Hashim
Department of Electrical and Electronic Engineering, Faculty of Engineering,
University Malaysia Sarawak (UNIMAS)
Kota Samarahan, Sarawak, Malaysia
hnizzati@unimas.my

Asrani Lit
Department of Electrical and Electronic Engineering, Faculty of Engineering,
University Malaysia Sarawak (UNIMAS)
Kota Samarahan, Sarawak, Malaysia
lasrani@unimas.my

2025 13th Asia-Pacific International Conference on Lightning (APL) | 979-8-3503-5694-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/APL65034.2025.11108819

***Abstract*—This study presents a novel thunderstorm prediction model leveraging a hybrid approach that integrates the Synthetic Minority Oversampling Technique (SMOTE), *k*-Means clustering and Machine Learning (ML) Models. Using historical lightning and meteorological data from the southern region of Peninsular Malaysia, the study evaluates the performance of five ML models including Decision Tree (DT), Random Forest (RF), Extra Trees (ET), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost), based on the standard performance evaluation metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that ensemble methods, particularly RF, consistently outperform other models across all three clusters, achieving prediction accuracy exceeding 95%. These findings underscore the effectiveness of RF in capturing data complexities and making accurate thunderstorm predictions. The study further emphasizes the role of balanced datasets through SMOTE and robust clustering techniques in enhancing model reliability. Future work will focus on integrating real-time data and incorporating additional meteorological variables to further improve predictive performance.**

***Keywords*—thunderstorm, lightning, machine learning (ML) models, clustering, synthetic minority oversampling technique (SMOTE), prediction model**

I. INTRODUCTION

Thunderstorms are among the most destructive natural phenomena, causing significant damage to infrastructure, loss of life, and disruptions to daily activities. Accurate prediction of thunderstorms is critical for mitigating their impact, particularly in regions like Peninsular Malaysia, which frequently experiences lightning due to its tropical, equatorial climate. The intense convective activity in this region contributes to a high occurrence of thunderstorms. Integrating historical lightning and meteorological data can provide valuable insights into the atmospheric precursors of such events. However, traditional prediction models often struggle in handling the nonlinear, high-dimensional nature of meteorological data, prompting the adoption of more advanced Machine Learning (ML) approaches. These limitations have motivated researchers to explore advanced ML techniques, which are better equipped to manage complex, high-dimensional datasets and uncover hidden patterns. Recent advancements in ML have demonstrated the potential of hybrid approaches that combine data pre-processing techniques, clustering methods, and ensemble learning models. One of the key challenges in meteorological datasets is class imbalance, where thunderstorm events represent a minority class. The Synthetic Minority Oversampling Technique (SMOTE) has proven effective in addressing this issue by synthetically generating samples for underrepresented classes during model training [1], [2]. In parallel, clustering techniques such as *k*-Means, enable the segmentation of data into meaningful groups, facilitating the development of specialized models tailored to specific data characteristics [3], [4]. Ensemble learning models, including DT, RF, ET, GB, and XGBoost have consistently outperformed traditional algorithms in various prediction tasks due to their ability to capture complex interactions and mitigate overfitting [5], [6], [7].

This study focuses on developing a thunderstorm prediction model for the southern region of peninsular Malaysia by integrating SMOTE, *k* -Means clustering, and multiple ML algorithms, including DT, RF, ET, GB, and XGBoost. The proposed work utilizes an extensive dataset comprising nine years of historical lightning data (2011–2019) from TNBR Malaysia combined with meteorological records provided by MetMalaysia. The analysis focused on a 100 km radius centered around the UTM-defined southern region of Peninsular Malaysia. Model performance is evaluated using standard performance evaluation metrics such as accuracy, precision, recall, and F1-score. By integrating SMOTE, *k*-Means clustering and ML techniques, this study aims to enhance the accuracy and reliability of thunderstorm predictions, contributing to improved disaster preparedness and risk management in the region.

The significance of this research lies in its potential to address the challenges associated with thunderstorm prediction in tropical climates, where traditional models often fall short. By incorporating clustering and sampling techniques, this study enhances both the accuracy and interpretability of predictions. The proposed model not only improves forecasting performance but also contributes to a deeper understanding of the underlying atmospheric dynamics. Moreover, it supports disaster preparedness and risk management efforts by offering a reliable predictive tool.

The remainder of the paper is organized as follows: Section II presents a comprehensive review of related work, highlighting recent advancements in thunderstorm prediction. Section III outlines the methodology, including data pre-processing, clustering, and model development. Section IV presents the results and discusses the performance of the proposed model. Finally, Section V concludes the study and suggests directions for future research.

II. LITERATURE SURVEY

A. Resampling Technique: Synthetic Minority Oversampling (SMOTE)

The Synthetic Minority Oversampling Technique (SMOTE) is a widely used data pre-processing method designed to address the issue of class imbalance in datasets by generating synthetic samples for the minority class. Introduced by Chawla et al., SMOTE generates synthetic samples for the minority class by interpolating between existing data points [2]. This approach has been extensively applied in various domains, including meteorological and storm prediction, where imbalanced datasets are common. For instance, studies have shown that SMOTE improves the performance of ML models in predicting rare weather events, such as thunderstorms, by ensuring that minority classes are adequately represented during training [1]. Moreover, recent advances such as Borderline-SMOTE and SMOTE-ENN offer further enhancements in handling noisy and overlapping data [3]. Borderline-SMOTE focuses on generating new samples near the decision boundary, particularly where the minority class is at higher risk of misclassification [8].

On the other hand, SMOTE-ENN combines SMOTE with data cleaning using the Edited Nearest Neighbours (ENN) method, which helps remove noisy or borderline majority instances [9]. These approaches have been shown to improve minority class recall and overall classification performance in imbalanced datasets [8], [9], making them promising alternatives for future improvement of thunderstorm prediction models. The integration of SMOTE with ensemble learning models has also been explored, demonstrating significant improvements in prediction accuracy and robustness [10].

B. Clustering Method: k-Means Clustering

k-Means clustering is a popular unsupervised learning algorithm used for partitioning datasets into distinct clusters based on similarity. Proposed by MacQueen, k-Means has been widely adopted in meteorological studies to segment data into meaningful groups, facilitating the development of specialized prediction models [11]. In the context of thunderstorm prediction, k-Means has been used to cluster the historical data into distinct regimes based on meteorological conditions (temperature, humidity, rainfall and wind speed) and lightning parameters (location, time, peak current, and polarity) into regions with similar weather patterns [4]. This approach seeks to improve the predictability of thunderstorm events by identifying both spatial and temporal trends. This clustering approach enables the creation of localized models that are better suited to the unique characteristics of each cluster. The primary advantage of integrating k-Means clustering with predictive modeling is the ability to extract salient features and patterns that can subsequently enhance the performance of supervised learning algorithms. Recent advancements in k-Means have further improved their efficiency and accuracy [4]. Additionally, hybrid approaches that combine k-Means with other clustering methods, such as hierarchical clustering, have been explored to enhance the quality of clusters and improve prediction performance.

C. Machine Learning (ML) Models

Machine learning (ML) models have revolutionized weather prediction by enabling the analysis of complex and high-dimensional data [12]. Among the various models, ensemble learning methods, such as DT, RF, ET, GB, and XGBoost, have gained prominence for their ability to capture nonlinear relationships and reduce overfitting.

The DT model is a simple yet powerful model that is widely used for its simplicity, interpretability, and ease of implementation, particularly in classification tasks like thunderstorm forecasting [13]. Building on this foundation, ensemble methods such as RF and ET combine multiple DTs to improve robustness and generalization. RF, introduced by Breiman, reduces variance by averaging predictions across multiple trees, while ET enhances this approach by introducing greater randomness during tree construction, which has shown improved performance in capturing complex atmospheric interactions [14], [15].

GB is a powerful ensemble method that builds models sequentially, optimizing for errors in previous iterations. GB models improve accuracy by sequentially correcting errors from previous models, making them suitable for modeling nonlinear dependencies in weather systems [16]. XGBoost, an advanced and efficient implementation of GB, has gained popularity for its scalability and regularization capabilities, and has demonstrated superior performance in thunderstorm prediction, particularly when used in conjunction with data pre-processing techniques like SMOTE [17].

D. Performance Evaluation Metrics

Performance metrics play a crucial role in evaluating the effectiveness of prediction models. In the context of thunderstorm prediction, standard performance evaluation metrics such as accuracy, precision, recall, and F1-score are commonly used [18], [19].

Accuracy measures the proportion of correctly predicted instances out of the total instances. While it provides a general overview of model performance, it may not be sufficient for imbalanced datasets, as it can be biased towards the majority class [19]. Precision quantifies the proportion of true positive predictions out of all positive predictions, indicating the model's ability to avoid false positives. It is particularly important in thunderstorm prediction, where false alarms can lead to unnecessary disruptions [19]. Recall, or sensitivity, measures the proportion of true positive predictions out of all actual positive instances. It is critical for ensuring that the model captures as many thunderstorms as possible, minimizing the risk of missed events. F1-score is the harmonic mean of precision and recall, providing a balanced measure of model performance. It is particularly useful for imbalanced datasets, as it accounts for both false positives and false negatives [19].

III. METHODOLOGY

To conduct this study, the methodology leverages a hybrid approach whereby the k -Means algorithm is applied first to partition the multidimensional dataset into clusters. Each cluster is then scrutinized for its characteristic atmospheric conditions, providing insights into pre-convective regimes [8]. The resulting cluster labels, along with the original features, serve as inputs for the ML models. This integrated strategy enables the extraction of crucial hidden variables that can boost classification performance. The prediction results are evaluated using standard performance metrics, including accuracy, precision, recall, and F1-score, ensuring a comprehensive assessment of the model's operational potential [9].

The methodology of this study is organized into four sections: A) Study Area and Data, B) Workflow, C) Clustering and ML Models, and D) Performance Evaluation Metrics and Hyperparameter Tuning. Each section is designed to systematically develop a robust framework for thunderstorm prediction using historical lightning and meteorological data.

A. Study Area and Data

The study area encompasses a 100 km radius centered on the UTM defined southern region of Peninsular Malaysia. This region is characterized by dense urban areas and frequent thunderstorms, combined with complex terrains, coastlines and monsoon-influenced climate patterns, contributing to unique convective dynamics. The spatial and temporal resolution of the data allows for an in-depth analysis of convective phenomena, making it well-suited for both clustering and classification tasks.

- Lightning Data (2011–2019): Sourced from TNBR Malaysia, covering cloud-to-ground lightning occurrences, which includes details on strike location (longitude, latitude), time, peak current (kA), and lightning polarity.
- Meteorological Data (2011–2019): Obtained from MetMalaysia, including temperature, relative humidity, wind speed, and atmospheric pressure.

B. Workflow

Fig. 1 depicts the workflow of the thunderstorm prediction model. The overall workflow starts with data collection and integration, wherein lightning and meteorological datasets are merged based on temporal and spatial parameters. Data cleaning, normalization and handling missing values are then performed. Outliers, such as abnormal peak currents and missing meteorological readings, are detected and handled using robust statistical techniques. Then, followed by the feature selection, where the identification of key meteorological parameters that influence thunderstorms. A resampling technique called SMOTE is introduced to balance the cluster data by identifying and connecting the nearest minority class data and creating synthetic data.

The clustering using k -Means is performed by grouping weather patterns to identify storm-prone conditions. After that, all the ML models are trained, such as DT, RF, ET, GB, and XGBoost. Lastly, the performance of the prediction model was evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score.

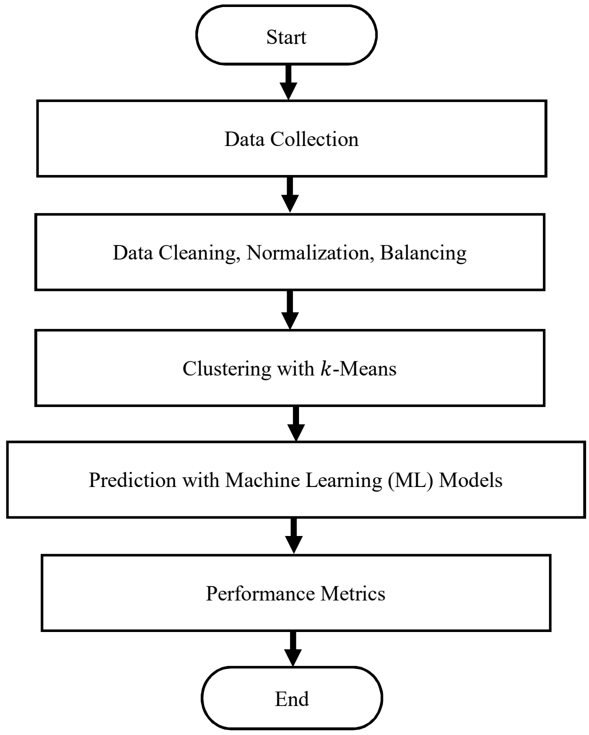


Fig. 1. Workflow for thunderstorm prediction model.

C. Clustering and Machine Learning (ML) Models

Clustering is an unsupervised ML technique used to group similar data points based on their features. In this study, k -Means clustering is applied to categorize different meteorological patterns that influence thunderstorm occurrences. k -Means clustering efficiently groups weather patterns with similar characteristics, which helps in distinguishing thunderstorm-prone conditions. k -Means clustering improves and reduce the complexity of meteorological data before feeding it into ML models. k -Means clustering has been successfully used in previous meteorological studies [3], [4], [7]. The k -Means clustering algorithm steps involve the following:

1. Initialization cluster center (centroids), where the k centroids is randomly selected.
2. Assigning each data point to the nearest cluster centroid based on the Euclidean distance, as in (1):

$$d(x,c) = \sqrt{\sum_{i=1}^{n}(x_i - c_i)^2} \tag{1}$$

Here, x represent the data point, and c represents the cluster centroid.

3. Recalculating the centroids as the mean of the assigned points.
4. The iteration is continued until convergence, where the centroid does not change significantly, and this indicate the clusters us stable.

5. To find the best number of clusters k, the elbow method is used, which plots the within-cluster sum of squares against different numbers of k clusters.

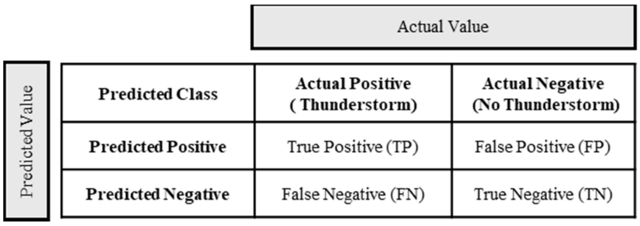
By applying k -Means clustering to meteorological and lightning data, this study categorizes distinct weather conditions that contribute to thunderstorms. These clusters then serve as input features for supervised ML models. The resampling technique called SMOTE is applied to overcome the imbalance in the clustered data by generating synthetic data. The balanced clustered data is then used to train and evaluate five ML models:

1. Decision Trees (DT): A tree-based model that splits the data based on feature thresholds. DT is a simple yet interpretable model.
2. Random Forest (RF): A bagging-based ensemble model that creates multiple DTs that reduce overfitting by averaging their predictions.
3. Extra Trees (ET): A variant of RF that uses random splits points instead of the best split for improved diversity.
4. Gradient Boosting (GB): An ensemble method that combines multiple weak models (DT) by iteratively improving performance by correcting previous errors.
5. Extreme Gradient Boosting (XGB): An optimized version of GB that improves speed and efficiency using parallel computation and regularization.

D. Performance Evaluation Metrics and Hyperparameter Tuning

The performance of the model to predict thunderstorm occurrences accurately is evaluated through a confusion matrix, also called as evaluation matrix. Fig. 2 displays the confusion matrix that is used to examine the entire set of models and emphasizes the frequency of the models produce correct or wrong predictions. The confusion matrix provides insights into four key components: the True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) rates, which are used to form the standard performance evaluation metrics such as accuracy, precision, recall, and F1-score. The TP and TN mean correctly predicted thunderstorms and correctly predicted no thunderstorms, respectively. While the FP represents false alarms for thunderstorms, and the FN means missed thunderstorms.

Table I displays the performance evaluation metrics used in the proposed work. Accuracy measures the proportion of correctly predicted instances out of all instances in the dataset. Precision indicates the proportion of true positives among all predicted positives. Recall represents the proportion of true positives among all actual positives. F1-score is the harmonic mean of precision and recall, providing a balanced measure of both.



Predicted Class	Actual Positive (Thunderstorm)	Actual Negative (No Thunderstorm)
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Fig. 2. Confusion matrix.

TABLE I. PERFORMANCE EVALUATION METRICS

Metric	Formula
Accuracy	$\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$
Precision	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$
Recall	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
F1-score	$2\,x\,\frac{\text{Precision } x \text{ Recall}}{\text{Precision }+\text{ Recall}}$

Noted that the range for all the metrics is within 0 to 1, and the optimum value is 1.

In addition to evaluating the model performance using a confusion matrix and standard performance evaluation metrics, this study incorporated hyperparameter tuning to further optimise the model performance.

Hyperparameters, which are configuration settings external to the model training process, significantly influence the learning behaviour and performance of ML algorithms. To identify the optimal hyperparameter values, techniques such as grid search and random search were employed in conjunction with cross-validation to systematically explore the best parameter combinations. This method was crucial in avoiding overfitting and improving the generalisation of the models.

As a result, the tuned models demonstrated enhanced accuracy, precision, recall, and F1-score, contributing to more robust and reliable thunderstorm predictions.

IV. RESULTS AND DISCUSSIONS

This work implements the k -Means clustering into the balanced dataset to find the optimal number of clusters. Fig. 3 shows the elbow curve of the optimum number of clusters for the datasets.

The elbow curve graph stabilises gradually after three clusters, indicating that the inclusion of additional clusters may impact accuracy. The balanced and clustered dataset is divided into 3 clusters, and the 3 coordinates of the centre for each cluster are identified, as shown in Fig. 4 and Table II, respectively. Cluster 1 has the largest proportion, followed by Cluster 2 and Cluster 3. This balanced clustering ensures that the dataset is well-distributed, which is crucial for accurate ML models training.

The clustering approach aligns with studies that emphasize the importance of balanced datasets for reducing bias in predictive models. Related works, such as those using k-Means clustering for weather prediction, have shown that balanced clusters improve the model performance metrics.

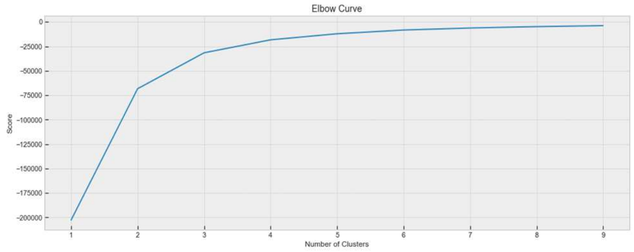


Fig. 3. The elbow curve.

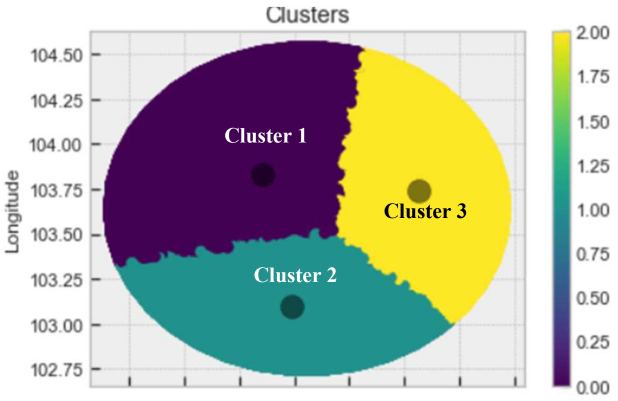


Fig. 4. Distribution of the balanced and clustered dataset.

TABLE II. COORDINATES OF THE CENTRE FOR EACH CLUSTER AND DATA COUNTS

	Coordinate		Total Count
	Latitude	*Longitude*	
Cluster 1	1.35519355	103.82931461	531785
Cluster 2	1.48598792	103.10133626	413961
Cluster 3	2.06637249	103.73642334	357478

The performance evaluation metrics of five ML prediction models at clusters 1, 2, and 3 are displayed in Fig. 5, Fig. 6, and Fig. 7, respectively. At cluster 1, all the prediction models performed well with accuracy, precision, recall and F1-score values exceeding 90.00% except for the F1-score of the ET model gave 70.53%, which is still acceptable. The high performance across all models suggests that the features used for prediction are highly informative. The superior performance of RF is consistent with its ability to handle imbalanced datasets and capture complex patterns, as highlighted in related studies [7], [13]. While SMOTE helped address the class imbalance in this study, recent oversampling techniques offer potential for further improvement. For instance, Borderline-SMOTE selectively generates synthetic samples near the decision boundary, focusing on minority instances most prone to misclassification [8]. Meanwhile, SMOTE-ENN combines oversampling with data cleaning using Edited Nearest Neighbours (ENN), which helps eliminate noisy or ambiguous samples from the majority class [9]. These advanced techniques have been shown to improve minority class recall and reduce overlap, and they are considered promising alternatives for future refinement of thunderstorm prediction models. In Cluster 2, all the prediction models gave an accuracy of 99.21% except the DT model gave 98.98% accuracy, which indicates a good prediction of thunderstorm at Cluster 2. The precision of all prediction models is 97.76% except for the DT model is 97.13%. The F1-score of ET, GB and XGBoost models outperforms the other models, followed closely by DT and RF. The consistent performance of all ML models across clusters demonstrates their robustness and generalizability. The results align with findings from other studies that highlight the effectiveness of ensemble methods like ET and RF in weather prediction tasks [7], [12], [13]. Differ from Cluster 3, all the performance metrics are slightly lower than those for Clusters 1 and 2 but still exceed 90%. The prediction models ranged from 90.00% to 91.00% in precision, but accuracy, recall, and F1-score gave more than 98.00%. RF remains the top-performing model, followed by ET and DT. The slight drop in performance could be due to differences in the characteristics of Cluster 3, such as fewer data points or higher variability. The results are consistent with studies that show ensemble methods like RF and ET excel in handling diverse weather datasets [5], [7]. Overall, RF demonstrates the effectiveness in predicting thunderstorms across different clusters, and the results are consistent with related works in the literature [4], [5], [7], [12], which emphasize the importance of balanced datasets and robust models for accurate weather prediction.

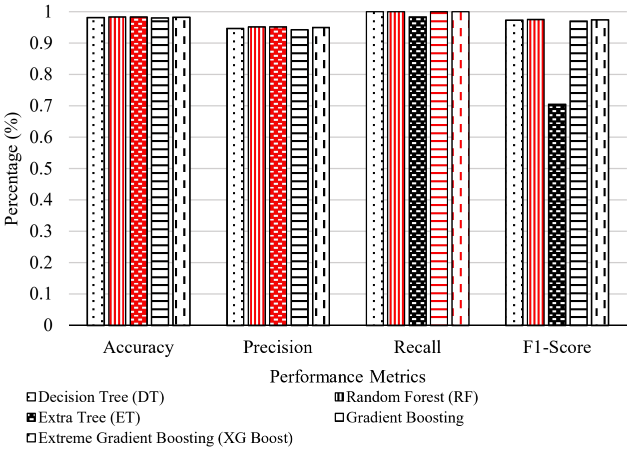


Fig. 5. Performance of the prediction model at Cluster 1.

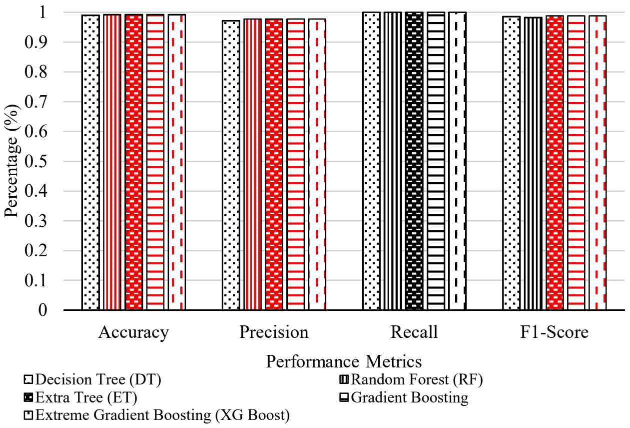


Fig. 6. Performance of the prediction model at Cluster 2.

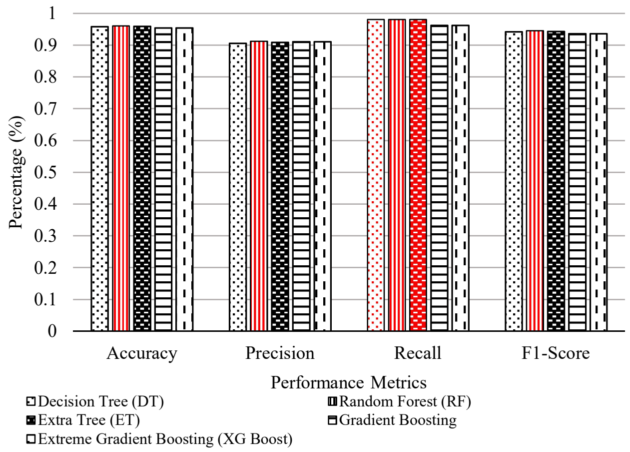


Fig. 7. Performance of the prediction model at Cluster 3.

V. CONCLUSION

This study introduced a hybrid thunderstorm prediction model integrating SMOTE, *k*-Means clustering, and ML algorithms to enhance predictive accuracy using historical lightning and meteorological data from the southern region of Peninsular Malaysia. The implementation of *k*-Means clustering ensured a balanced dataset, which significantly improved the reliability of ML models in thunderstorm forecasting. The models are evaluated using standard performance evaluation metrics: accuracy, precision, recall and F1-score. Among the five ML models evaluated—Decision Tree (DT), Random Forest (RF), Extra Trees (ET), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost)—the Random Forest (RF) model consistently outperformed others, achieving over 95% accuracy across all clusters. The balanced clustering approach ensures unbiased model performance, while the high metrics across all clusters validate the robustness of the selected features and models. The findings confirm that ensemble methods, particularly RF, effectively capture complex weather patterns, making them suitable for operational thunderstorm prediction. Additionally, the study highlights the critical role of balanced datasets and clustering techniques in improving predictive model performance. The use of alternative oversampling methods such as Borderline-SMOTE and SMOTE-ENN will be investigated to further improve model performance in handling class imbalance. These approaches offer targeted improvements by refining decision boundaries and removing noisy data points, thereby enhancing prediction accuracy and reliability. Future research will focus on integrating real-time meteorological data to enhance the model's responsiveness to evolving weather conditions. This integration aims to enable timely updates and maintain predictive accuracy in dynamic environments. Implementing adaptive models capable of processing streaming data with low latency will be essential for delivering prompt and reliable predictions. Addressing challenges such as data quality, system scalability, and model stability will be crucial to ensure the effectiveness of real-time thunderstorm prediction systems.

ACKNOWLEDGMENT

The authors would like to thank the Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM), and the Research, Innovation and Enterprise Centre (RIEC), Universiti Malaysia Sarawak (UNIMAS), for the financial and management support. The cooperation from lightning technical expert, Ir. Noradlina Abdullah and other technical staff of TNBR is highly appreciated. This research uses data provided by the Lightning Detection System Network (LDSN), managed by the Lightning Detection System Laboratory, TNB Research Sdn. Bhd. (TNBR), and meteorological data from MetMalaysia for research and educational purposes.

REFERENCES

[1] A. Fernández, S. García, F. Herrera, and N. V Chawla, 'SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary', 2018.

[2] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 'SMOTE: Synthetic Minority Over-sampling Technique', 2002.

[3] D. Arthur and S. Vassilvitskii, '*k*-Means++: The Advantages of Careful Seeding'.

[4] A. Yovan Felix, G. S. S. Vinay, and G. Akhik, '*k*-Means Cluster Using Rainfall and Storm Prediction in Machine Learning Technique', *J Comput Theor Nanosci*, vol. 16, no. 8, pp. 3265–3269, Aug. 2019, doi: 10.1166/jctn.2019.8174.

[5] O. O. A., Z. I. A., and A. L., 'Comparative Analysis of Weather Prediction Using Classification Algorithm: Random Forest Classifier, Decision Tree Classifier and Extra Tree Classifier', *African Journal of Mathematics and Statistics Studies*, vol. 7, no. 2, pp. 162–171, May 2024, doi: 10.52589/ajmss-f6h03bne.

[6] T. Chen and C. Guestrin, 'XGBoost: A scalable tree boosting system', in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[7] R. Meenal, P. A. Michael, D. Pamela, and E. Rajasekaran, 'Weather prediction using random forest machine learning model', *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, pp. 1208–1215, May 2021, doi: 10.11591/ijeecs.v22.i2.pp1208-1215.

[8] H. Han, W.-Y. Wang, and B.-H. Mao, 'Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning', in *LNCS*, vol. 3644, 2005, pp. 878–887. doi: 10.1007/11538059_91.

[9] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, 'A study of the behavior of several methods for balancing machine learning training data', *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.

[10] H. Han, W.-Y. Wang, and B.-H. Mao, 'LNCS 3644 - Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning', 2005.

[11] J. Macqueen, 'Some Methods for Classification and Analysis of Multivariate Observations'.

[12] H. Zhang, Y. Liu, C. Zhang, and N. Li, 'Machine Learning Methods for Weather Forecasting: A Survey', Jan. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/atmos16010082.

[13] O. O. A., Z. I. A., and A. L., 'Comparative Analysis of Weather Prediction Using Classification Algorithm: Random Forest Classifier, Decision Tree Classifier and Extra Tree Classifier', *African Journal of Mathematics and Statistics Studies*, vol. 7, no. 2, pp. 162–171, May 2024, doi: 10.52589/ajmss-f6h03bne.

[14] L. Breiman, 'Random Forests', vol. 45, pp. 5–32, Apr. 2001.

[15] P. Geurts, D. Ernst, and L. Wehenkel, 'Extremely randomized trees', *Mach Learn*, vol. 63, no. 1, pp. 3–42, Apr. 2006, doi: 10.1007/s10994-006-6226-1.

[16] J. H. Friedman, 'Greedy Function Approximation: A Gradient Boosting Machine', *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001, [Online]. Available: https://www.jstor.org/stable/2699986

[17] S. Guastavino *et al.*, 'Prediction of severe thunderstorm events with ensemble deep learning and radar data', *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-23306-6.

[18] H. M and S. M.N, 'A Review on Evaluation Metrics for Data Classification Evaluations', *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 01–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201.

[19] David M W Powers, 'Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation', 2007.