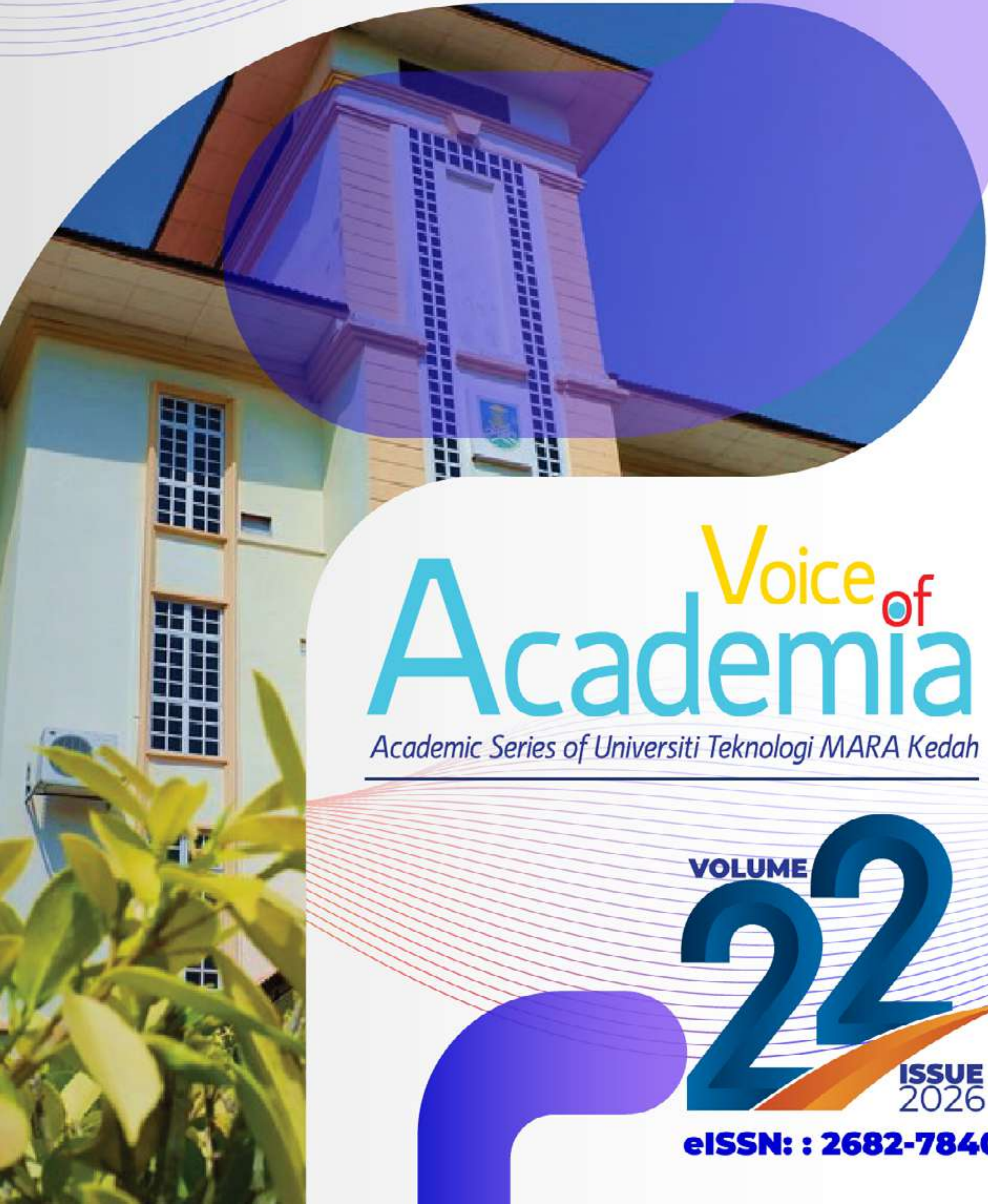
UNIVERSITI
TEKNOLOGI
MARA



Voice of Academia

Academic Series of Universiti Teknologi MARA Kedah

VOLUME 22
ISSUE 1
2026

eISSN: : 2682-7840

DR. IRA PATRIANI,
FISIP, UNIVERSITAS TANJUNGPURA (UNTAN),
PONTIANAK, INDONESIA

DR. RIZAL ZAMANI IDRIS,
FACULTY OF SOCIAL SCIENCE & HUMANITIES,
UNIVERSITI MALAYSIA SABAH (UMS), SABAH

DR. SIMON JACKSON,
FACULTY OF HEALTH, ARTS AND DESIGN,
SWINBURNE UNIVERSITY OF TECHNOLOGY MELBOURNE, AUST

DR. AZYYATI ANUAR,
FACULTY OF BUSINESS MANAGEMENT,
UNIVERSITI TEKNOLOGI MARA (UITM) KEDAH BRANCH, MALAYSIA

DR. FARYNA MOHD KHALIS,
COLLEGE OF CREATIVE ARTS,
UNIVERSITI TEKNOLOGI MARA (UITM) SHAH ALAM, MALAYSIA

DR IDA NORMAYA MOHD NASIR,
FACULTY COMPUTER SCIENCE AND MATHEMATICS,
UNIVERSITI TEKNOLOGI MARA (UITM) KEDAH BRANCH, MALAYSIA

DR MOHD FAIZAL JAMALUDIN,
FACULTY OF ACCOUNTANCY,
UNIVERSITI TEKNOLOGI MARA (UITM) KEDAH BRANCH, MALAYSIA

DR. MUHAMAD KHAIRUL ANUAR ZULKEPLI,
ACADEMY OF LANGUAGE STUDIES,
UNIVERSITI TEKNOLOGI MARA (UITM) KEDAH BRANCH, MALAYSIA

DR NOR ARDIYANTI AHMAD,
FACULTY OF ADMINISTRATIVE SCIENCES & POLICY STUDIES,
UNIVERSITI TEKNOLOGI MARA (UITM) KEDAH BRANCH, MALAYSIA

e-ISSN:2682-7840

9 772682 784004

© Voice of Academia is jointly published by the Universiti Teknologi MARA Caawangan Kedah, Malaysia and Penerbit UiTM (UiTM Press), Universiti Teknologi MARA Malaysia, Shah Alam, Selangor.

TABLE of CONTENTS

FORECASTING THE MALAYSIAN RINGGIT (MYR) EXCHANGE RATE: ARIMA VS GARCH 1 - 16
Rabi'atul Adawiyah Muhamad Shah & Nurul Nisa' Khairol Azmi*

BEHIND THE SCREEN: A SYSTEMATIC REVIEW OF CONTEMPORARY CHALLENGES IN DIGITAL LEARNING 17 - 36
Siti Noorsiah Jamaludin[1*], Abd Samad Hasan Basari[2]

INVESTIGATING THE EFFECTIVENESS OF COLLABORATIVE LEARNING STRATEGIES IN MASTERING THE ARABIC LANGUAGE 37 - 52
Abd Rahman Jamaan*

IMPLEMENTING CLAY SCULPTING AS AN IDEATION STRATEGY IN TEACHING PRODUCT FORM DESIGN TO FIRST-YEAR INDUSTRIAL DESIGN STUDENTS 53 - 68
Mohd Hamidi Adha Mohd Amin[1*]

EXPLAINING ENTREPRENEURIAL INTENTION OF MALAYSIAN PUBLIC UNIVERSITY STUDENTS: THE MEDIATION MODERATED MODEL 69 - 88
Ahmad Nabil Mohd Zahariman [1], Nur Fatin Syazliana Zahar [2], Nurul Hidayana Mohd Noor[3*] & Syeliya Md Zaini[4]

SUPPLIER SELECTION OF HALAL KOREAN RESTAURANT USING FUZZY TOPSIS 89 - 102
Norpah Mahat[1*], Nurul Aqilah Ahmad [2]

RAINFALL INTENSITY CLASSIFICATION IN SUBANG 103 - 116
Mohamad Aiman Hakim Mohamad Nizam[1], Isnewati Ab Malek[2*] & Jaida Najihah Jamidin[3]

SYNCHRONOUS AND ASYNCHRONOUS CORRECTIVE FEEDBACK FOR GRAMMAR ACCURACY: ESL NOVICE TEACHERS' BELIEFS AND PRACTICES 117 - 130
Aiman Zulaikha Mohd Fadzli [1], Sheela Faizura Nik Fauzi[2*] & Abdul Azim Mahda[3]

DIGITAL SPORTS GRAPHICS AND BRAND PERSONALITY IN THE MALAYSIAN SEPAK TAKRAW LEAGUE 131 - 142
Muhammad Asyraf Hanafi[1], Neesa Ameera Mohamed Salim[2*] & Azhar Abd Jamil[3]

TOWARDS INCLUSIVE GAMIFIED LITERACY INTERFACES FOR MALAYSIAN STUDENTS WITH DYSLEXIA: INTEGRATING THE DELONE AND MCLEAN INFORMATION SYSTEMS SUCCESS MODEL 143 - 163
Safura Adeela Sukiman[1*]

UNDERSTANDING RECYCLING BEHAVIOR: A STUDY ON UITM SEGAMAT STUDENTS 164 - 176
Nur Diana Zaman[1], Fatin Farazh Ya'acob[2*], Basri Badyalina[1], Muhammad Zulqarnain Hakim Bin Abd Jalal[1],Amir Imran Zainoddin[2] & Kerk Lee Chang[1]

DOES FDI BENEFIT ALL? EXAMINING INCOME INEQUALITY ACROSS 10 ASEAN NATIONS 177 - 190
Bee-Hoong Tay[1*], Nurulrahwani Hamsan[2], Nurul Dhihani Md Idris[3] & Nurul Hafiizzati M Roslee[4]

MESOPOTAMIAN ARCHITECTURE AS BACKGROUND DESIGN IN CONTEMPORARY ANIMATED SERIES 191 - 209
Nureen Qistina Affandi[1], Siti Nur Ain Abd Rahman[2*]

TRANSFORMING HRM EDUCATION THROUGH VALUES-BASED PEDAGOGY: THE ADAB+ APPROACH AND CORPORATE RELEVANCE **210 - 224**
Muhammad Aiman Awalluddin[1*], Anisa Safiah Maznorbalia[2] &
Mohd Ramlan Mohd Arshad[3]

COMPARATIVE ANALYSIS OF PSEUDOCODE AND FLOWCHARTS IN ALGORITHM DEVELOPMENT AMONG FIRST-YEAR COMPUTER SCIENCE STUDENTS **225 - 239**
Satria Arjuna bin Julaihi[1], Zubaidah binti Bohari[2*], Rumaizah binti Che Md Nor[3]
& Abdul Hadi bin Abdul Talip[4]

EXPLORING THE STUDENT PERCEPTION OF ACRONYM-BASED LEARNING APPROACH IN LEARNING ACCOUNTING PRINCIPLES AMONG NON-ACCOUNTING MAJOR STUDENTS **240 - 258**
Siti Aimi Mohamad Yasin[1] , Corina Joseph[2*] & Nur Izyan Ismail[3],
Nuraisyah Fitrie Abdullah@Abd Jalil[4], Azmira Abdullah[5]

ZAKAT DISTRIBUTION DECISION BASED ON FUZZY EVALUATION APPROACH **259 - 272**
Zamali Tarmudi[1], Noor Syazana Ngarisan[2*] & Muhammad Yassar Yusri[3]

LEAN PRACTICES IN CONSTRUCTION: A COMPREHENSIVE LITERATURE REVIEW ON ENHANCING PROJECT PERFORMANCE **273 - 284**
Syed Nasrul Fadzli Syed Mohamad[1*], Mohd Shahnaz Bin Mahbook Ali[2] &
Amir Ahzlina Jasni[3]

Voice of Academia Vol.22 Issue (1) 2026

Voice of Academia

e-ISSN: 2682-7840

RAINFALL INTENSITY CLASSIFICATION IN SUBANG: A DECISION TREE APPROACH

Mohamad Aiman Hakim Mohamad Nizam[1], Isnewati Ab Malek[2]* & Jaida Najihah Jamidin[3]

[1,2,3] Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Seremban,
70300 Seremban, Negeri Sembilan, Malaysia

ARTICLE INFO

Article history:

Received Aug 2025
Accepted Oct 2025
Published Jan 2026

Keywords:

Rainfall intensity, Decision Tree, Gini, Entropy & Logworth

Corresponding Author:
isnewati@uitm.edu.my

ABSTRACT

Subang, Malaysia, is increasingly vulnerable to flash floods due to its urban development and frequent weather disturbances. Accurate rainfall forecasting is therefore essential for mitigating flood risks and supporting effective urban planning. This study employs Decision Tree Analysis to classify rainfall intensity into two categories: rain and no rain. Historical meteorological data, including temperature, humidity, wind speed, and mean sea level pressure, were analyzed to identify the key environmental determinants of rainfall. Three decision tree models, based on Gini, Entropy, and Logworth indices were applied to classify rainfall events and evaluate predictive performance. The analysis reveals that temperature and relative humidity exert the strongest influence on rainfall intensity, followed by mean sea level pressure. Among the models, the Gini index demonstrated superior accuracy in detecting rainfall events, whereas the Entropy-based model provided the most consistent performance in terms of generalization and sensitivity to unseen data. These findings highlight the capacity of decision tree methods to capture nonlinear interactions within meteorological variables while producing interpretable and practical forecasts. The study underscores the potential of such models to enhance flood management strategies and urban resilience. Furthermore, it recommends continuous data enrichment and the integration of additional meteorological variables to improve prediction accuracy.

1. Introduction

Rainfall is not only a vital component of the Earth's hydrological cycle but also a key factor in managing water resources, sustaining agriculture, regulating climate, and preserving ecosystems (Yoga et al., 2022). Understanding rainfall intensity, the amount of rainfall over a specific period, typically measured in millimeters per hour or day, is crucial for accurate weather forecasting, especially in regions vulnerable to extreme climatic events. Rainfall intensity directly influences water availability, crop productivity, urban planning, and disaster preparedness. It also plays a critical role in flood prediction, irrigation scheduling, and drought mitigation, issues that have grown more pressing as climate change has intensified the frequency and severity of extreme weather events (Mamun et al., 2018).

In Malaysia, rainfall patterns are highly variable and unpredictable, particularly in urbanized states such as Selangor and Kuala Lumpur. These areas often experience flash floods during the Southwest Monsoon (May–September) and Northeast Monsoon (November–March), when rainfall events are heavy, localized, and capable of overwhelming drainage systems (Tugi et al., 2023). Such events disrupt transportation, damage infrastructure, and lead to significant economic losses, as highlighted by Tanzizi (2025), who reported severe flooding in Selangor that rendered roads impassable and caused widespread traffic congestion. The ability to classify rainfall intensity whether slight, moderate, heavy, or very heavy is therefore essential for effective disaster management, climate adaptation, and sustainable urban development (Tella et al., 2023).

The challenge is further compounded by climate change, which has increased rainfall variability and made extreme events more common, posing one of Malaysia's greatest threats to water resource management (Mignot & Dewals, 2022). Heavy downpours often trigger flash floods in urban centers like Subang, where rapid urbanization, dense populations, and limited drainage capacity heighten flood risks. These conditions highlight the urgent need for reliable rainfall prediction models to strengthen urban resilience and mitigate the adverse effects of climate change.

Recent advances in machine learning provide promising tools for this purpose. Models such as decision trees, artificial neural networks, and multiple linear regression have been successfully applied to predict rainfall patterns in different regions (Liyew & Melese, 2021). Decision trees offer a transparent and interpretable approach to classification, enabling urban planners to better design stormwater systems, optimize water storage, and develop effective flood control measures. By identifying key variables such as temperature, humidity, and mean sea level pressure, decision tree models can help forecast rainfall events with greater accuracy, ensuring more informed decision-making.

Ultimately, this study aims to classify rainfall intensity in Subang using decision tree models, focusing on rain and no rain categories. Subang is especially vulnerable to flash floods due to its high level of urbanization and frequent extreme rainfall events. By improving rainfall prediction in such areas, the findings may contribute to more sustainable water management, stronger disaster preparedness, and greater resilience in the face of climate change.

2. Literature Review

Understanding rainfall intensity and its prediction requires consideration of multiple climatic variables, particularly in tropical countries such as Malaysia, where urban areas like

Subang are highly vulnerable to flash floods. Rainfall is influenced by interconnected atmospheric conditions, including temperature, humidity, mean sea level pressure (MSLP), and wind speed, all of which affect the likelihood, timing, and severity of precipitation. Recent studies have also highlighted the effectiveness of machine learning models, particularly decision tree algorithms, in capturing nonlinear relationships among these variables and improving rainfall classification accuracy (Raniprima et al., 2024; Samadianfard et al., 2022).

Temperature is one of the most critical drivers of rainfall variability. Both minimum and maximum temperatures influence atmospheric moisture content and convective activity. Tan et al. (2020) reported that Malaysia experienced significant warming trends between 1985 and 2018, with minimum temperatures increasing by 0.27°C and maximum temperatures by 0.12°C per decade. This warming, exacerbated by the urban heat island effect in cities like Subang, has led to hotter nights and intensified rainfall events. Empirical evidence suggests that incorporating temperature variables into machine learning models enhances predictive accuracy. Raniprima et al. (2024), for example, demonstrated that decision tree and random forest models using temperature, humidity, and wind direction achieved accuracies as high as 95.64%. Similarly, Suhaila et al. (2011) confirmed that temperature fluctuations strongly influence rainfall patterns across Malaysian regions.

Relative humidity also plays a vital role in rainfall formation. High atmospheric moisture increases the probability of cloud development and precipitation. Studies in Malaysia have shown that humidity levels in Subang typically range between 70% and 90%, with rainfall becoming highly probable when humidity exceeds 80% (Idris, 2024; Noor et al., 2023). Decision tree models often incorporate humidity as a predictor variable, as it closely correlates with rainfall intensity and provides reliable thresholds for classifying events (Lenderink et al., 2025).

In addition, MSLP is a key factor in tropical rainfall systems. Fluctuations in pressure are linked to larger-scale phenomena such as the monsoon and the Intertropical Convergence Zone (ITCZ). Research indicates that a decrease in MSLP often signals incoming heavy rainfall, as low-pressure systems draw in moisture and trigger storm development (Clayson et al., 2019; Nikumbh et al., 2021). Machine learning studies have successfully applied MSLP as an input variable to classify rainfall categories in Selangor, although predictability is improved when combined with other variables (Hussin et al., 2025; Marufuzzaman et al., 2022).

Wind speed is another important determinant of rainfall intensity. Higher wind speeds are typically associated with convective storms and heavy precipitation events, particularly during Malaysia's monsoon seasons (Swarno et al., 2020). Gonzalez et al. (2022) and Gao et al. (2021) further emphasized that wind speed enhances the ability to forecast both the occurrence and severity of rainfall when combined with humidity and temperature in decision tree models.

Rainfall intensity itself serves as the dependent variable in classification models, often categorized into light, moderate, and heavy rainfall, though some studies simplify the classification into binary outcomes such as rain and no rain (Rosmadi et al., 2023). Previous research has demonstrated that decision trees can effectively capture threshold-based interactions between climatic variables. For instance, conditions such as temperature above 30°C, humidity greater than 80%, and high wind speed often predict heavy rainfall, while lower values correspond to light or no rainfall (Wardani et al., 2023; Yap & Jamaludin, 2024).

Collectively, these findings highlight the importance of multiple meteorological factors in rainfall prediction and the growing utility of machine learning models, particularly decision trees, in handling complex, nonlinear interactions among variables. By integrating temperature, humidity, MSLP, and wind speed, decision tree models provide interpretable and accurate classifications, making them valuable tools for flood risk management and urban planning in climate-vulnerable regions such as Subang.

3. Methodology

This study adopts a quantitative research design to classify rainfall intensity in Subang, Selangor, using machine learning approaches. Historical meteorological data were collected, encompassing variables such as minimum and maximum temperatures, relative humidity, mean sea-level pressure, and wind speed.

Rainfall intensity served as the dependent variable, categorized into two classes based on daily rainfall amount: rain (> 0 mm) and no rain (0 mm). This threshold is consistent with standard meteorological practice, where any measurable precipitation greater than zero is considered a rainfall event. While rainfall is traditionally classified into multiple categories (e.g., light, moderate, heavy), simplifying it into a binary classification makes the model more practical and interpretable for machine learning applications. Oswal (2019) demonstrated that such binary classification is particularly effective when paired with decision tree algorithms, which thrive on clear distinctions between classes.

Study Area

The focus of this research is Subang, Selangor, an area that has increasingly faced the challenges of urban flooding due to rapid urbanisation and changing rainfall patterns. Subang holds a strategic position within Selangor's urban and economic network, serving as a key transportation and industrial hub. It is surrounded by residential, commercial, and industrial developments, including major roads, Subang Airport, and several industrial factories. These features highlight the socio-economic importance of Subang while also underlining its vulnerability to extreme weather events.

The choice of Subang as a case study area is strongly supported by its history of flood events. For instance, in March 2022, Subang experienced flash floods that caused significant disruptions to traffic and local activities (Bernama, 2022). Such occurrences reflect the inadequacy of existing drainage systems in coping with heavy rainfall and illustrate the urgency of developing reliable predictive models for rainfall intensity. By focusing on Subang, this research situates itself in a context where the results can have direct implications for flood preparedness and urban planning in a densely populated, high-risk urban environment.

Data Sources

The meteorological data used in this study were obtained from two reliable institutions: the Malaysian Meteorological Department (MetMalaysia) and the Department of Irrigation and Drainage (DID). These agencies provide official weather and hydrological datasets that are frequently used for academic and applied research in Malaysia.

The dataset spans ten years, from January 2014 to December 2024, offering a comprehensive representation of rainfall variability in Subang. It includes daily precipitation measurements as well as supporting meteorological variables such as temperature, humidity, mean sea level pressure, and wind speed. This extended dataset length ensures that the analysis captures both seasonal and inter-annual variations in rainfall, including patterns associated with Malaysia's two main monsoon seasons: the Southwest Monsoon (May–September) and the Northeast Monsoon (November–March).

By incorporating data across an entire decade, the study not only captures regular climatic cycles but also reflects anomalies and extreme events, which are becoming more frequent with climate change. Such coverage strengthens the robustness of the predictive models developed, as they are trained and tested on a wide variety of weather conditions.

Method of Analysis

The primary analytical approach adopted in this study is Decision Tree Analysis, a supervised machine learning algorithm widely used for both classification and regression tasks. Decision trees were selected due to their ability to handle nonlinear relationships among variables and their inherent interpretability, which makes them suitable for practical decision-making in environmental management.

Decision trees function by recursively partitioning the dataset into subsets based on the most informative predictor variables. At each node, a decision rule is applied to split the data, continuing until terminal nodes (leaves) represent the classification outcomes in this case, "rain" or "no rain." Unlike traditional statistical models that assume linearity, decision trees can capture complex interactions among variables such as temperature, humidity, and pressure without requiring prior assumptions about their distribution.

Before model development, several data preprocessing steps were conducted. Missing values were identified and handled using appropriate imputation techniques to ensure data completeness. Outliers were examined to prevent extreme values from disproportionately influencing the decision tree splits.

For this study, the dataset was divided into training and testing subsets in a 70:30 ratio. The training set was used to build the decision tree, while the testing set evaluated the model's generalisation ability on unseen data. This approach ensures that the model is not only fitted to past data but is also capable of producing accurate forecasts under new conditions. Figure 1 shows a flowchart of the process for generating the decision tree analysis and obtaining the study's results.

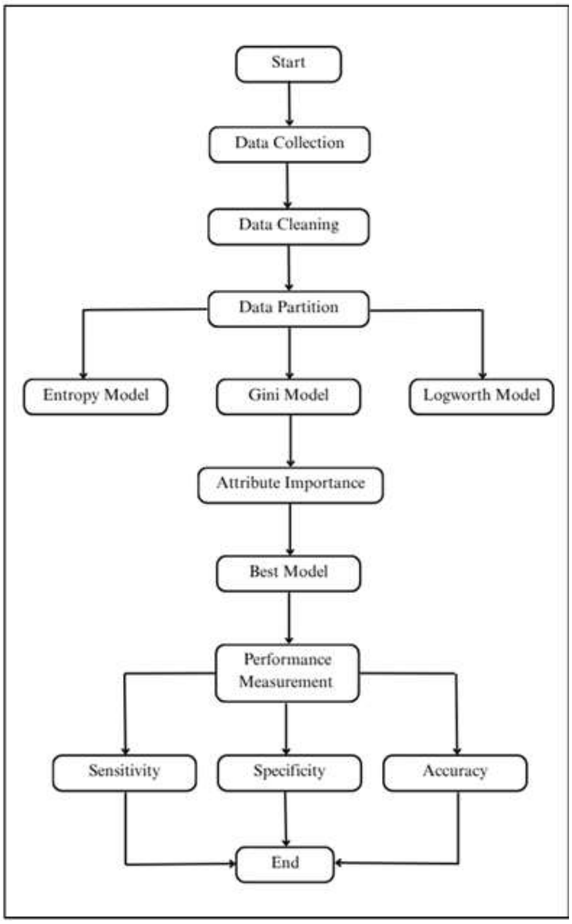


Figure 1. Flowchart of Decision Tree

Gini Index – Measures the impurity of a dataset, with lower values indicating more homogeneous splits. It prioritises variables that most effectively reduce uncertainty in rainfall classification.

Entropy – Derived from information theory, entropy quantifies disorder or randomness in data. A lower entropy indicates more reliable splits. This measure aims to maximise information gain at each step of the decision tree.

Logworth – Provides a statistical measure of a variable's significance based on p-values, with higher values indicating stronger predictive importance. This criterion complements Gini and Entropy by highlighting the most influential variables in rainfall prediction.

These criteria were used to evaluate which environmental variables best contributed to predicting rainfall and to determine which decision tree model provided the most reliable classification. The analysis also incorporated attribute importance measures, which rank variables based on their contribution to the model's predictive power. This step directly addressed one of the study's objectives: identifying the key meteorological variables influencing rainfall in Subang.

Model performance was assessed using standard classification metrics: accuracy, sensitivity, specificity, and misclassification rate. Accuracy measured the overall correctness of predictions, sensitivity captured the model's ability to correctly detect rainfall events, and specificity reflected its ability to correctly identify non-rain days. By comparing these metrics across the Gini, Entropy, and Logworth models, the study identified the most effective method for rainfall classification in the Subang context.

The methodology was designed to align closely with the study's two main objectives: (1) determining the important meteorological variables influencing rainfall intensity in Subang, and (2) identifying a reliable method for classifying rainfall using decision tree algorithms. Through attribute importance analysis and model evaluation, the study systematically addressed these objectives.

4. Results

This chapter explained and clarified the results derived from the examination of the rainfall intensity in Subang by using a decision tree, namely Gini, Entropy, and Logworth. This study aims to identify a reliable method for classifying rainfall intensity, as well as to determine the important variables in predicting the rainfall intensity in Subang.

Table 1 presents the relative importance of each attribute in the three-decision model which are Gini, Entropy and Logworth. The model considered five meteorologist attributes, namely minimum temperature, maximum temperature, mean relative humidity, mean sea level pressure (MSLP) and mean wind speed. However, not all attributes are equally distributed across all models.

The attributable proved that decision tree models mainly focused on attributes based on the influence of the target variables. The splitting criteria applied in the decision tree (Entropy, Gini, Logworth) generated importance values for each of the five attributes that were utilized. Mean Relative Humidity is an example of a high-value attribute that facilitated more effective splits. The exclusion of other attributes during the splitting process indicates that not all attributes significantly in the process of developing a reliable model.

Table 1
Attributes Importance for Logworth, Gini and Entropy Models

Attributes	***Gini***	***Entropy***	***Logworth***
Mean Relative Humidity	1.0000	1.0000	1.0000
Maximum Temperature	0.1556	0.1209	0.1427
Minimum Temperature	0.1524	0.1146	0.0000
Mean Sea Level	0.0749	0.0000	0.0000
Mean Wind Speed	0.0798	0.0000	0.0000

Table 1 shows that the significant attributes across all models remain the same, specifically Mean Relative Humidity and Maximum Temperature. According to Daniya et al. (2020), the Gini coefficient is a measure to assess the impurity and purity of a dataset, based on how pure the data is, thereby selecting the optimal model for data partitioning at each node within the decision tree. Lower Gini values indicate more homogeneous classifications.

On the other hand, Entropy measures the level of uncertainty in classifying the rainfall intensity indices, based on the chosen variables. A low value of entropy means that the data is more predictable. Logworth, on the other hand, is a method used in the decision trees to evaluate and see the importance of a certain model. This indicates that a higher Logworth value indicates that the split is more significant and potentially leading to a better model performance.

To achieve the first objective of this study, which is to identify the important variables for classifying the rainfall intensity in Subang, this objective can be achieved through variable importance analysis across the three decision tree models. The results of the analysis hold significant value as they highlight all the key factors that help in creating the decision tree model. As shown in Table 1, the variables or attributes of Mean Relative Humidity show the highest significance value which is up to 100% or 1.0000. This shows that within all three models, Mean Relative Humidity holds the most important variables.

Table 2 shows the fit statistics for three different decision tree models: Gini, Entropy, and Logworth. Each model is assessed using several metrics, including the misclassification rate, average squared error, and ROC index, both for training and validation sets, as well as the gaps between these values.

Table 2
Fit Statistics of Decision Tree Model

Model Description	Gini	Entropy	Logworth
Valid: Misclassification Rate	0.2433	0.2452	0.2462
Train: Misclassification Rate	0.2157	0.2202	0.2206
GAP	0.0276	0.0249	0.0255
Valid: Average Squared Error	0.1647	0.1696	0.1799
Train: Average Squared Error	0.1523	0.1566	0.1664
GAP	0.0124	0.0129	0.0134
Valid: ROC Index	0.8260	0.8110	0.7620
Train: ROC Index	0.8460	0.8350	0.7870
GAP	-0.0200	-0.0240	-0.0250

To evaluate model performance, the differences (gaps) between training and validation results were examined to identify underfitting and overfitting among the three decision tree models: Gini, Entropy, and Logworth. Underfitting was assessed by inspecting the gaps for Average Squared Error (ASE), Misclassification Rate (MCR), and ROC Index. An underfit model would typically exhibit negative gaps for ASE and MCR and a positive gap for the ROC Index. Based on this criterion, none of the models showed clear evidence of underfitting.

Overfitting was identified by comparing the magnitude of the gaps between training and validation performance for ASE, MCR, and ROC Index, with the sign of the ROC gap disregarded. Among the three models, the Gini model exhibited the largest gaps, indicating a higher risk of overfitting compared to the Entropy and Logworth models. Consequently, the Gini model was classified as overfitted and excluded from consideration for the final model.

The remaining models, Entropy and Logworth, were then compared based on their validation performance. The Entropy model achieved a lower validation ASE (0.1696) and MCR (0.2452), along with a higher validation ROC Index (0.8110), compared to the Logworth model. These results indicate that the Entropy model provides a better balance between predictive accuracy and generalization performance. Therefore, the Entropy model was selected as the best-performing model overall.

Although the Gini model demonstrated superior raw performance metrics, including the lowest validation MCR (0.2433), lowest validation ASE (0.1647), and highest validation ROC Index (0.8260), its larger gaps between training and validation results suggest reduced generalizability. While it achieved the highest classification accuracy and strongest discriminative ability, the evidence of overfitting limits its suitability for deployment on unseen data. The ROC Index further supports these findings. The Gini model recorded the highest ROC values on both the training (0.8460) and validation (0.8260) sets, followed by the Entropy and Logworth models. Despite this, the Entropy model demonstrated more stable performance across datasets, reflecting better generalization capability.

In conclusion, although all three models show acceptable performance for rainfall classification, the Entropy-based decision tree is identified as the most suitable model due to its balanced accuracy, lower error measures, and reduced overfitting risk. This aligns with existing literature, which highlights both Gini and Entropy criteria as reliable for classification tasks, with Entropy often demonstrating improved generalization performance.

Sensitivity, specificity, and accuracy are three primary measures for assessing the predictive performances of models (Table 3). Sensitivity, which is also referred to as the true positive rate, is a measure how effectively a model can detect positive cases. It can be calculated by dividing the number of true positives by the sum of the true positives and the false negatives. The Gini model demonstrates the highest sensitivity, indicating superior performance in identifying positive cases. In contrast, specificity considers how well a model can detect negative cases. It is calculated as the true negatives divided by the sum of the true negatives and the false positives. The Entropy model appears to excel in this regard; it provides the highest specificity value, which implies that it can make negative predictions very well. Finally, accuracy indicates the overall performance of the model, in terms of correctness within all the positive and negative cases combined. Accuracy formula considers true positives, true negatives, false positives and false negatives. According to Table 3, the Gini model achieves the best performance in accuracy and represents the most robust model. Each of these measures emphasises a different aspect of the model assessment: the Gini model excels in sensitivity and accuracy, while the Entropy model demonstrates superior specificity.

Table 3
Sensitivity, Specificity and Accuracy for All Decision Tree Model

Model Description	***Sensitivity***	***Specificity***	***Accuracy***
Gini	0.8270	0.6488	0.7567
Entropy	0.8190	0.6561	0.7548
Logworth	0.8206	0.6512	0.7538

For the original model based on Gini, it has the Sensitivity of 0.8270, which says that the model has correctly predicted 82.70% of the time when it rained (Positive Class). This high sensitivity indicates that the model can detect rainfall events well. In particular, the Gini model has a Gini value of 0.6488 equal to a minimum rate with it correctly detects cases with no rain (i.e., the negative class). And even those areas with high AUC have only a moderate level of the same it is a sign that the Gini model is less effective at predicting no-rain cases than rain cases. The Gini model has a total accuracy of 0.7567, which means 75.67% of all predictions the model makes, whether it is going to rain or not, are correct. This makes the Gini a robust model, particularly for the detection of rain, but shows limitations regarding its performance capability for no-rain conditions.

The sensitivity of the Entropy model (0.8190) is not as high as Gini model's but remains adequate for predicting rain events. Its specificity is 0.6561, lower than Gini models, since the Entropy model is worse at predicting the Breach nonevent. The accuracy of the Entropy model is 0.7548, very close to the Gini model's, indicating both models perform equally in overall predictions.

Finally, the sensitivity of the Logworth model is 0.8206, very close to the Gini and Entropy models, indicating adequate performance in rain prediction. Its specificity is 0.6512, which is slightly higher than the Gini model's and clearly it is lower than the Entropy model. The accuracy of the Logworth model is 0.7538, which is slightly lower than the other two models (Gini: NTippe-Entropy & Gini), but it is good enough to make a difference that the Logworth model is reliable for overall performance.

Overall, the three models generate comparable results, and the differences in sensitivity, specificity, and accuracy are not large. The Gini model demonstrates a slight advantage in sensitivity or how well it predicts rain, while the Entropy model shows marginally superior specificity. Although slightly more accurate than the Gini and Entropy models, the Logworth model is a suitable alternative when a balance between sensitivity and specificity is required.

5. Discussion

All three models (Gini, Entropy, and Logworth) have shown that Relative Humidity holds the most important variable in classifying rainfall intensity, as noted by Khosravi et al. (2025), who show that Relative Humidity serves as a crucial indicator for classifying rainfall intensity. It measures the amount of moisture in the atmosphere, which affects the rainfall patterns and intensity. Moreover, it is important to know that atmospheric conditions as well as the Relative Humidity, can make an impact on the accuracy of rainfall predictions, thereby affecting the weather forecasting and climate studies.

The meteorological variable mean sea level pressure (MSLP) might show its importance across all decision tree models due to its correlation with other variables. According to the research by Lee et al. (2024), MSLP is highly correlated with all temperature variables. This can happen especially in countries or regions where temperature is a main factor in atmospheric pressure changes. In these cases, MSLP does not provide additional information beyond what is already captured by variables like temperature or humidity. This could lead to lower significance of its variable across all three models. Furthermore, MSLP may not be directly linked to rainfall intensity because other variables such as Mean Relative Humidity exhibit a more immediate and direct relationship with rainfall occurrence (Allan et al., 2015).

Since no importance values were generated during the analysis, the Mean Wind Speed may also not be significant across all models, as shown in Table 1. This lack of importance may be explained by the rainfall intensity classification in Subang, the selected area, where wind speed may not have a significant impact on rainfall patterns. Van De Walle et al. (2021) agreed with this by claiming that the wind speed determines the energy of the storm but not the intensity of the rainfall that comes directly from it if the weather is not stormy. Then, the main cause of storm dynamics is the wind speed. In areas where the rainfall intensity is more of a function of such elements as air moisture (humidity), the wind speed may be less relevant to model accuracy. Furthermore, Yang and Tsai (2019) found that although wind speed is crucial in the occurrence of extreme weather events, it is not the main factor in predicting moderate or regular rainfall. Therefore, wind speed may demonstrate lower importance in the decision tree models.

6. Conclusion

The results of this study provide strong evidence that decision tree models, particularly Gini, Entropy, and Logworth models, are effective for classifying the rainfall intensity in the Subang area. The tree-based models not only are able to easily forecast the magnitude of rainfall intensity based on limited meteorological variables but also provide clear and understandable outputs that are essential for practical concerns such as urban planning and disaster management. Relative humidity was the most important variable of the tested three models in all decision trees. This finding aligns with the literature, which emphasizes the importance of humidity in determining rainfall intensity. All three decision tree models identified relative humidity as a major driver of precipitation, with relative humidity proving to be critical to the successful classification of rainfall.

The Gini model's performance especially regarding sensitivity and accuracy, is high, making it particularly useful for identifying rainfall events. However, the Entropy model, which exhibited slightly higher specificity, demonstrated superior air temperature prediction. It showed stronger generalization to unseen data, important for real-world applications which operate in settings when future data may differ from the training data. It is apparent that, although on the training data Gini performs slightly better than the results from Entropy, for the transfer stage, Entropy is more reliable for rainfall intensity forecasting than the Gini model. While the Logworth model performed well, it did not outperform the Gini and Entropy models in terms of classification accuracy and reliable prediction.

The findings from the study highlight the usefulness of decision tree models to aid local planners and policy makers in making effective decisions. Assessing rainfall intensity levels with decision trees can be especially useful in flood risk mitigation measures, as it may influence better forecasting of rainfall events and consequent impacts on urban infrastructure. Now this becomes more relevant in urbanized areas such as Subang where the urbanization is fast and the drainage systems are not properly designed to cope with flash floods. In addition, the necessity of further investigation with respect to other meteorological factors was underscored such as wind speed, mean sea level atmospheric pressure, as well as other atmospheric conditions for improvement of the further developed rainfall intensity classification models.

In summary, decision tree models found to be a useful and efficient approach for classifying the intensity of rainfall. They offer the advantage of high interpretability as well as strong predictive accuracy, which are important for decision-making on urban flood risk management. By identifying the significant meteorological parameters, these models provide valuable insights into the complex characteristics of rainfall patterns and the mechanisms influencing flood risk in urban settings.

Acknowledgments

The authors are grateful to Universiti Teknologi MARA (UiTM), Seremban branch, Malaysian Meteorological Department (MetMalaysia) and the Department of Irrigation and Drainage (DID) for providing the data. They also extend their gratitude to other researchers for their valuable ideas and discussions.

Authors Contributions

Mohamad Aiman Hakim Mohamad Nizam is a full-time Bachelor of Science (Hons.) in Statistics student and was responsible for data analysis and writing the original draft of the manuscript. Isnewati Ab Malek, Senior Lecturer, contributed through supervision, as well as reviewing and

editing the manuscript. Jaida Najihah Jamidin, Senior Lecturer, contributed by reviewing the manuscript and revising the methodology and formatting.

Conflict of Interest

The authors declare there is no conflict of interest in the subject matter or materials discussed in this manuscript.

DOI

References

Allan, R. P., Lavers, D. A., & Champion, A. J. (2016). Diagnosing links between atmospheric moisture and extreme daily precipitation over the UK. *International Journal of Climatology*, *36*(9), 3191-3206.

Bernama. (2022, March). Flash floods hit Subang, major disruptions reported. Bernama News Agency.

Clayson, C. A., Roberts, J. B., & Robertson, F. R. (2019). Sea surface fluxes associated with precipitation extremes. Journal of Climate, 32(10), 2899–2917. doi:10.1175/JCLI-D-18-0587.1

Daniya, T., Geetha, M., & Kumar, K. S. (2020). Classification and regression trees with gini index. *Advances in Mathematics: Scientific Journal*, *9*(10), 8237-8247.

Gao, L., Liu, Y., & Wang, S. (2021). Impacts of wind speed on precipitation intensity in tropical regions. Atmospheric Research, 253, 105322. doi:10.1016/j.atmosres.2021.105322

Gonzalez, P., Rivera, D., & Morales, J. (2022). Wind speed as a predictor for extreme rainfall in Southeast Asia. Climate Dynamics, 59, 345–362. doi:10.1007/s00382-021-05931-4

Hussin, M. H., Rahman, N. H. A., & Yusof, N. (2025). Application of mean sea level pressure in rainfall classification models for Selangor, Malaysia. Theoretical and Applied Climatology, 150, 77–89.

Idris, N. (2024). Relative humidity and its influence on rainfall intensity in Selangor. Malaysian Journal of Climate Studies, 12(1), 23–34.

Kenawy, A. M., Lopez-Moreno, J. I., & Vicente-Serrano, S. M. (2019). Temperature extremes and their impact on rainfall variability in tropical climates. International Journal of Climatology, 39, 2587–2605. doi:10.1002/joc.5972

Khan, M., Awan, S., & Zafar, M. (2024). Decision tree models for rainfall prediction: A case study in South Asia. Environmental Modelling & Software, 165, 105665.

Khosravi, K., Farooque, A. A., Bateni, S. M., Jun, C., & Dhiman, J. (2025). Prediction of three vital rainfall characteristics: Advanced hybrid tree-or lazy-based learner?. *Results in Engineering*, *25*, 103840.

Lee, T., Won, C. H., & Singh, V. P. (2024). ML-based regionalization of climate variables to forecast seasonal precipitation for water resources management. *Machine Learning: Science and Technology*, *5*(1), 015019.

Lenderink, G., Barbero, R., Loriaux, J. M., & Fowler, H. J. (2025). Relative humidity and extreme rainfall in tropical climates. Nature Climate Change, 15, 245–252.

Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, *8*(1), 153.

Mamun, A. A., Chowdhury, S., & Alam, M. (2018). Rainfall variability and its impact on water resources in Southeast Asia. Environmental Earth Sciences, 77, 334. doi:10.1007/s12665-018-7499-4

Marufuzzaman, M., Ali, M., & Rahman, M. (2022). Improving rainfall predictability with pressure-based models in Southeast Asia. Atmospheric Science Letters, 23, e1114. doi:10.1002/asl.1114

Mignot, E., & Dewals, B. (2022). Urban flooding: Challenges for drainage and flood protection systems. Water, 14(7), 1103. doi:10.3390/w14071103

Nikumbh, A., Pai, D., & Mandal, G. S. (2021). The role of mean sea level pressure in monsoon rainfall variability. Climate Dynamics, 56, 2783–2800. doi:10.1007/s00382-020-05628-y

Noor, N. F. M., Hassan, Z., & Rahman, N. A. (2023). Classification of rainfall intensity using machine learning techniques in Malaysia. Asia-Pacific Journal of Atmospheric Sciences, 59, 567–582.

Oswal, S. (2019). Binary rainfall classification using machine learning models. International Journal of Artificial Intelligence Research, 6(2), 145–154.

Patel, A., Singh, R., & Gupta, P. (2023). Extreme rainfall and urban flooding: Implications for tropical cities. Natural Hazards, 117, 123–140. doi:10.1007/s11069-023-05982-7

Pizzorni, P., Russo, S., & Messori, G. (2022). Climate change and the rising variability of rainfall extremes. Environmental Research Letters, 17, 045013. doi:10.1088/1748-9326/ac5c7e

Raniprima, R., Widodo, P., & Rahmawati, N. (2024). Application of decision tree and random forest algorithms in rainfall prediction. Journal of Hydrometeorology, 25, 233–245.

Ridwan, M., Abdullah, S., & Karim, M. (2020). Comparative analysis of machine learning models for rainfall prediction in Malaysia. Malaysian Journal of Environmental Management, 21(3), 45–60.

Rosmadi, F., Aziz, A. A., & Ismail, M. R. (2023). Rainfall intensity classification for urban planning in Malaysia. Journal of Environmental Hydrology, 31, 101–113.

Samadianfard, S., Amiri, M., & Mosavi, A. (2022). Decision tree-based approaches for rainfall prediction. Water Resources Management, 36, 4571–4592.

Sharma, A., Singh, R., & Kumar, S. (2021). Machine learning approaches in rainfall prediction: A review and case study. Applied Computing and Geosciences, 11, 100072.

Suhaila, J., Deni, S. M., & Zin, W. Z. W. (2011). Relationship between temperature fluctuations and rainfall in Malaysia using Functional Data Analysis. Theoretical and Applied Climatology, 103, 405–416.

Swarno, A., Latif, M. T., & Yusop, Z. (2020). Monsoon-related wind speed and its association with heavy rainfall in Selangor. Meteorological Applications, 27, e1943.

Tan, M. L., Samat, N., Chan, N. W., & Lee, J. (2020). Temperature extremes and trends in Malaysia from 1985 to 2018. Atmospheric Research, 234, 104739. doi:10.1016/j.atmosres.2019.104739

Tanzizi, A. (2025, January 15). Heavy rainfall floods Selangor roads, massive congestion reported. The Star. Retrieved from https://www.thestar.com.my

Tella, A., Yusuf, I., & Adeyemi, O. (2023). Rainfall intensity classification and its role in disaster management. International Journal of Climatology, 43, 1124–1137. doi:10.1002/joc.7932

Tugi, C. M., Aziz, N. A., & Hamid, S. (2023). Seasonal rainfall patterns during Southwest and Northeast Monsoon in Peninsular Malaysia. Malaysian Journal of Science, 42, 89–101.

Van de Walle, J., Thiery, W., Brogli, R., Martius, O., Zscheischler, J., & van Lipzig, N. P. (2021). Future intensification of precipitation and wind gust associated thunderstorms over Lake Victoria. *Weather and Climate Extremes*, *34*, 100391.

Wardani, S., Putra, R., & Nugroho, D. (2023). Machine learning classification of rainfall intensity in Southeast Asia. Environmental Modelling & Software, 158, 105489.

Yang, T. H., & Tsai, C. C. (2019). Using numerical weather model outputs to forecast wind gusts during typhoons. *Journal of Wind Engineering and Industrial Aerodynamics*, *188*, 247-259.

Yap, W. C., & Jamaludin, N. (2024). Predicting rainfall in tropical regions using temperature, humidity, and wind speed. Journal of Tropical Meteorology, 30, 56–70.

Yoga, G. P., Ismail, N. F., & Rahman, M. (2022). Rainfall and ecosystem sustainability: Impacts on climate change adaptation. Sustainability, 14(12), 6789. doi:10.3390/su14126789

اونيۏرسيتي تيكنولوݢي مارا
UNIVERSITI
TEKNOLOGI
MARA

9 772682 784004

eISSN: : 2682-7840