Journal of Sustainability Science and Management
Volume 18 Number 7, July 2023: 88-103

eISSN: 2672-7226
© Penerbit UMT

RAINFALL INTENSITY CLASSIFICATION IN THE EAST COAST OF MALAYSIA USING DISCRIMINANT ANALYSIS

MOHAMAD AMEER IMRAN MOHD NOOR, MUHAMMAD AFIQ HALEK, AZWAN FAIZ LIM MUHAMMAD RAZWAN LIM AND HASFAZILAH AHMAT*

Faculty of Computer and Mathematical Sciences, Kompleks Al-Khawarizmi, Universiti Teknologi MARA, 40000 Shah Alam, Selangor, Malaysia.

**Corresponding author: hasfazilah@uitm.edu.my*

Submitted final draft: 26 February 2023 *Accepted: 15 April 2023* http://doi.org/10.46754/jssm.2023.07.006

Abstract: In the previous study, principal component analysis and cluster analysis were used but no information on factors, contribution and classification for rainfall were provided. The logistic regression was not suitable for the rainfall classification since it only works well if the target variable is in binary output. This paper discusses the classification of rainfall based on the contribution of several factors, namely temperature, humidity, wind direction and wind speed on the east coast of Peninsular Malaysia using discriminant analysis. The trend of rainfall intensity was also identified using diurnal variation and Mann Kendall trend test. This study used the data from 2018 to 2020, which covered three locations on the east coast region; Kuala Krai (Kelantan), Kuala Terengganu (Terengganu), and Temerloh (Pahang) furnished by the Malaysian Meteorological Department. There were significant positive relationships among all independent variables, namely, temperature, humidity, wind direction and wind speed, with the rainfall intensity with the significant *p*-value of Wilk's Lambda <0.05. The findings indicated that the classification equation differs from location to location due to different levels of rainfall intensity, the location of monitoring stations and the factors affecting rainfall in these locations.

Keywords: Discriminant analysis, rainfall, classification, diurnal variation, trend.

Introduction

Malaysia is in a region at low risk of severe natural disasters. Earthquakes, typhoons, or volcanic eruptions have not affected the country. However, Malaysia is not spared from disasters like floods, landslides or haze (Ahmad *et al.*, 2017). The climate in Peninsular Malaysia is described as an equatorial climate with a latitude of 1° 15′ – 6° 45′ N and a longitude of 100° –104° 30′ E (Khan *et al.*, 2019). The topography can be described as mountains or highlands in the interior, with flat coastal zones (Wong *et al.*, 2016; Ziarh *et al.*, 2021). Stretching over 130,598 km^2, the region is distinguished by a high level of humidity, uniform temperature, and an extensive amount of rainfall (Khan *et al.*, 2019). Due to the geographical location of the country, the humidity throughout the year exceeds 68% (Mayowa *et al.*, 2015; Wong *et al.*, 2009; 2016). The temperature of the country differs according to the regions and time, where it varies between 23°C to 32°C with a mean of 27°C. The weather of the country is described as hot and humid due to its location, temperature, and humidity. According to Muhammad *et al.* (2019), the weather is affected by the two monsoons - Northeast and Southwest - that hit the region annually, bringing with it heavy rain. As a tropical country, Malaysia is among the top countries that have high rainfall. Flood occurrence is very common and it does not exclusively affect Malaysia as other countries like Indonesia and Vietnam also had their own experiences in facing this unprecedented event. Annually, the number of days that rainfall occurs in a year ranges from 150 days to 200 days (Muhammad *et al.*, 2019). Many researchers believe that prolonged heavy rain contributes to the occurrence of floods on the east coast (Ismail & Haghroosta, 2018; Maisarah *et al.*, 2019; Tam *et al.*, 2019). Historically, Malaysia is is prone to floods (Aziz *et al.*, 2016). Alias *et al.* (2016) stated that Malaysia is in the middle rank of the index of countries with the highest risk of natural disaster, occurrence which includes floods, storms, and other common natural disasters.

Globally, flood is said to be the most frequent natural disaster, where it damages the infrastructure and disrupts socioeconomic activities of a location, and even causes deaths (Ahmad *et al.*, 2017; Mignot *et al.*, 2019). Typically, two types of floods usually occur in Malaysia; flash floods and monsoonal floods (Ahmad *et al.*, 2017). The causal of floods is based on various factors, either natural or manmade factors. As for the natural factor, Malaysia experiences monsoon season where episodes of heavy or extreme rainfall can last up to seven days somewhere from November until January. Heavy rainfall can lead to severe floods in a certain region or place (Tam *et al.*, 2019). Nevertheless, the cause of floods may differ from one place to another. Some places experience flooding due to human activities that ignore the environmental issues and climate change in that area (Arief Rosyidie, 2013). In a study of the Epidemiology of Disasters (CRED) conducted by the Belgium-based Centre for Research, the biggest problem faced by Malaysia in the past two decades was flooding. This claim was supported by the Emergency Events Database (EM-DAT), where Malaysia faced fifty-one (51) natural disasters from 1998 to 2018, and floods logged the most cases every year. In this regard, floods affected many parties especially the public, resulting in 770,000 people being evacuated, 148 deaths, and causing roughly RM5.82 billion in losses for the last two decades. Most of the major flood events were associated with the northeast monsoon, which causes heavy rain on the east coast of Peninsular Malaysia (Alias *et al.*, 2016).

In many research papers, the extreme flood in 2014 that hit three states on east coast was unprecedented in the history of Malaysia. Hashim *et al.*, (2021) deduced that the east coast was the hardest hit by the flood. Two phases of rainfall triggered the flood to be among the worst. Elfithri *et al.* (2017) however, stated that the big flood affected all districts in Pahang at different time frames. Districts like Kuantan, Pekan, Raub, and Bentong were flooded around December to January. With 68,000 flood victims, the Department of Irrigation and Drainage (DID) estimated losses to be about RM73 million. During the same year, Terengganu also experienced big flood which forced more than 68,000 people to evacuate their houses. Terengganu encountered 3 waves of flood and suffered approximately RM12 million in damage (Aziz *et al.*, 2016). Even so, the same researcher believed that the floods that hit Kelantan in the same year was the worst among the three states.

The extreme flood in Kelantan was called the Kelantan Big Yellow Flood of 2014. The name 'Yellow' comes from the colour of the water containing sediment from the catchment area (an area into which the rainfall flows. It is recorded that the flood reached up to 5 to 10 meters, which is equivalent to the 3rd or 4th floor of a building. More than 150,000 people were hit by the floodand many of them could not evacuate their homes due to high-rise water. A large number of people could not get help as their supplies of food and other necessities ran out (Alias *et al.*, 2016; Ismail & Haghroosta, 2018). As a result, the Malaysian government had to allocate approximately RM2.8 billion for flood recovery (Tam *et al.*, 2019). The flood cause major damage, as illustrated in Figure 1.

In recent years, extreme rainfall events in Malaysia have become more common and affected many parties (Muhammad *et al.*, 2021). From various studies predicted that rainfall trends will continue to increase and lead to heavier rain on the east coast of Peninsular Malaysia. The need to identify the trend of rainfall is important to ensure that water resources are well-protected for future planning and development which includes flood control, agricultural activities, and drought management. However, all these studies still lacked focus on rainfall events on the east coast of Malaysia, which can be regarded as the most at-risk climate change zone in Malaysia (Mayowa *et al.*, 2015).

There are several statistical analyses and modelling to predict the intensity of rainfall by

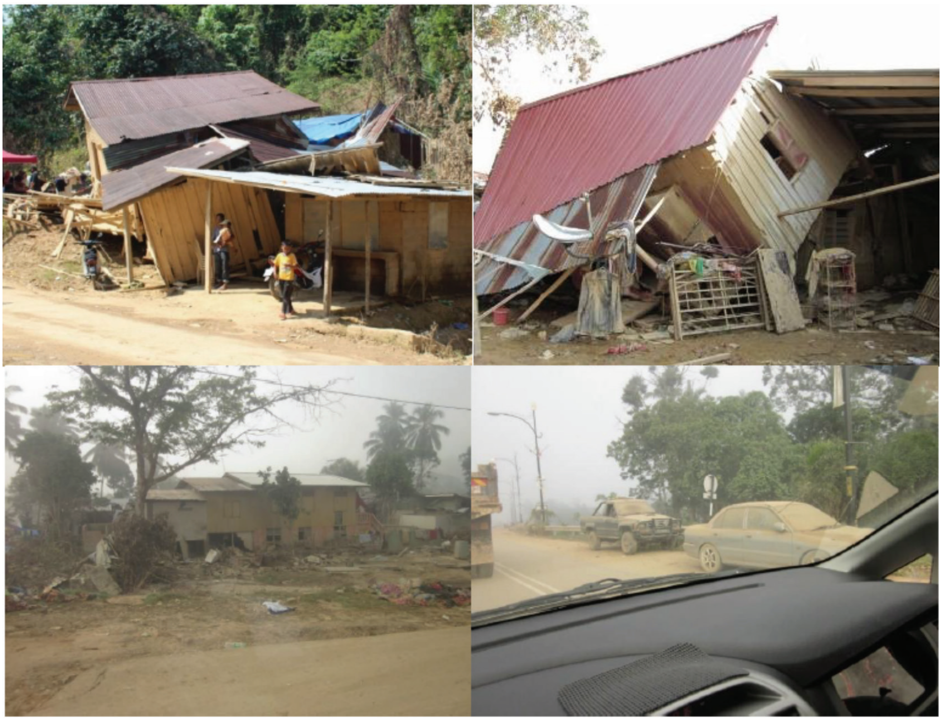


Figure 1: The aftermath of the Kelantan Big Yellow Flood of 2014 (self-collection)

previous researchers. The statistical methods used involve multiple linear regression, extreme value distributions (EVD), probability analysis, predictive analysis, and time series analysis. As for the EVD, the distributions used to model rainfall intensity involve exponential, beta and pareto distributions. The ARIMA model was commonly used in time series data to predict the behavioural rainfall pattern in the future. The logistic regression method holds the same approach to classify the dependent variable, however, the approach of logistic regression limits classification only into two groups. In this regard, discriminant analysis is the preferred method as it can classify more than two groups and identify the factors that contribute to the intensity of rainfall. It is important to understand whether meteorological or hydrological factors contribute the most to the intensity of rainfall. .

Various research has classified rainfall data (Mohd Ariff *et al*., 2019; Shaharudin & Ahmad, 2019; Sulaiman *et al*., 2020; Abu Bakar *et al.*, 2020) using Principal Component Analysis, Support Vector Machine or time series clustering such as Ward's Hierarchical Clustering. However, there are still large knowledge gaps in the study of the intensity of rainfall using discriminant analysis. The discriminant analysis provides a good statistical method to help the researcher check whether there are significant differences among groups and various sectors that have employed this analysis, such as the marketing sector distinguishing the factors that contribute to the preference of the customer, weather classification either dry or wet, and others. Although other research used discriminant analysis in the prediction of floods, it did not take into account the factors of rainfall or classify the rainfall intensity. In addition, there are limited studies predicting the level of rainfall intensity, either high or low. Discriminant analysis helps researchers determine which predictor variables are associated with other dependent variables since it has multiple benefits and advantages, similar to regression analysis (Keskin *et al.*, 2020). The purpose of this paper is to recommend the classification of rainfall intensity using discriminant analysis. It distinguishes the level of rainfall intensity using meteorological factors (humidity, location, temperature, wind speed, and wind direction) and observed significant relationships among the variables that can lead to floods.

Materials and Methods

Scope of Study

This study utilized hourly data of five meteorological parameters: rainfall (mm), temperature (°C), humidity (%), wind speed (m/s) and wind direction (°) from 2018–2020 furnished by the Malaysian Meteorological Department. Three locations - Kuala Krai (Kelantan), Kuala Terengganu (Terengganu), and Temerloh (Pahang) - were selected due to them being highly affected by rainfall and part of the flood-prone area on the east coast of Peninsular Malaysia compared to the other locations.

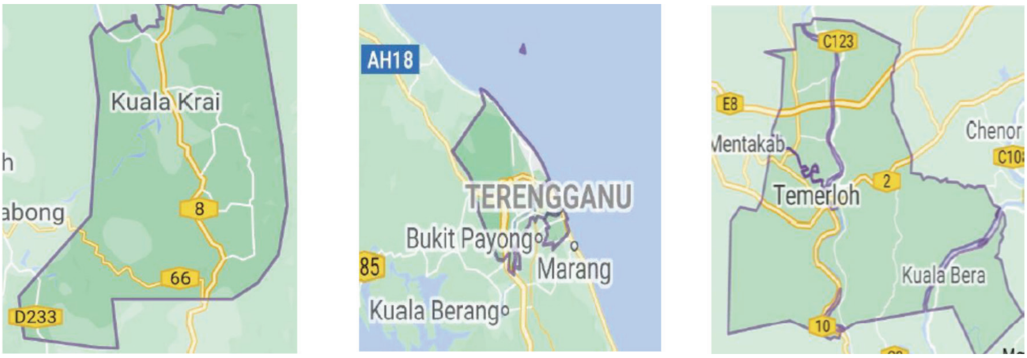
The Kuala Krai Station is in the centre of Kelantan at a latitude of 5.5308°N and a longitude of 102.2019°E. Kuala Krai covers an area of around 2,287 km^2 and has a population of 109,461. In 2014, Kelantan, specifically Kuala Krai, faced a huge flood known as the Big Yellow Flood (Alias *et al*., 2016). Kuala Terengganu is on the east coast of Peninsular Malaysia, and situated at the estuary of the Terengganu River and covers a total area of 605 km^2. It is at a longitude 5.3296°N and latitude 103.1370°E. It has a population of 426,500. In addition, Kuala Terengganu is near the beach and receives heavy rainfall every month. Temerloh is in central Pahang at longitude 3.4486°N and latitude 102.4163°E. It covers around 2,251 km^2 and has a population of 158,724. Temerloh is also known as the centre of Peninsular Malaysia and the second-largest town in Pahang. Temerloh was highly affected by the 2014 flood, according to Elfithri *et al.* (2017). The map of these three locations is illustrated in Figure 2.

Data Pre-processing

The data was checked for any anomalies in terms of missing values and outliers.

Missing Value Treatments

There are various reasons why missing values in meteorological data occur, i.e., a malfunction of equipment, human error, and calibration process. Complete data are required for performing trend analysis in this study. Generally, missing values can affect the effectiveness or accuracy of the result. In various research related to rainfall, missing values were often excluded since the researchers were not keen to check the trend. When a missing value was excluded, it lead to biases reduced the capability to detect a pattern (Ridwan *et al.*, 2021). There are several methods for treating the missing values, and one of the most popular approaches to s is by replacing them with any measure of central tendency, be it mean, mode or median. Mean replacement



(a) (b) (c)

Figure 2: Map of (a) Kuala Krai, Kelantan, (b) Kuala Terengganu, Terengganu, and (c) Temerloh, Pahang

is by calculating the mean of the available data and substituting it into the missing value. Mode on the other hand is the observation with the highest frequency, while the median is the middle value of the dataset. These measures can be used since the value would not affect the normality of the dataset. Since the focus of this study was to find the trend of rainfall on the east coast of Malaysia, the missing values need to be treated. In a recent study by Nor *et al.* (2020), the researcher concluded that replacing the missing value with the mean was the best method to treat the problem of missing rainfall data in east coast Peninsular Malaysia. Hence, this study adopted this method to treat the problem of missing values.

Outliers' Treatment

Another anomaly in any dataset was the outliers. The outlier is a data point that differs significantly from other observations in the dataset: whether it is extremely low or extremely high. For each location, the outlier was checked using the Mahalanobis distance test, since it involves multivariate analysis. First, the Mahalanobis distance was calculated using SPSS and then computed the cumulative distribution function for a chi-square distribution with 4-degree freedom. Then, the value will be compared with $\alpha = 0.001$ (Brereton, 2015) and if the distance value is less than 0.001, it indicates that the set of variables contained an outlier. If more than 50% of the Mahalanobis distance value falls below the chi-square or critical value, the data is normally distributed. For these three locations, there were several outliers detected from the total of 26,304 data for each location. Outliers can especially affect multivariate data when dealing with the equation. In this research, the researcher uses a box plot to check for the presence of outliers. An outlier may also be the error when the imputation data process is done. According to Leys *et al.* (2019) when it is not possible to retrieve the correct value, outliers should be deleted. Therefore, this study omitted extreme values since the dataset obtained was large.

Data Partitioning

The dataset was split into 2 sets: the training and validation datasets. The training set is used for learning and implementation in building a model. This dataset was used to fit the data to the discriminant analysis and check the performance of the model. A validation set, in contrast, is used to validate the model after the model was obtained using the training set. The validation is to verify the performance of the model. The proportion of splitting was 70%; 30% for the training and validation dataset. It is the most general proportion to be used as the idea is to make the model classification better in the training set since the error estimate is more accurate for validation data.

Transformation

Several transformations were applied to the data in terms of classification of the variables. There were 5 variables, including 1 dependent and 4 independent variables, used in this study. The data were transformed into values to meet the research objectives. As shown in Table 1, the parameter transformations were the wind direction (°) and rainfall amount originally in the unit of (mm) which are quantitative variables. In applying discriminant analysis, the variable rainfall amount which is the dependent variable should be a categorical variable. Hence, the rainfall amount variable was classified into 3 categories, which are low, moderate, and heavy rainfall intensity (in one hour) (Department of Irrigation and Drainage, n.d.).

Data Exploration

Descriptive Statistic

The descriptive statistic was used to understand the characteristic of all meteorological parameters in this study. The analysis includes mean, standard deviation (SD), minimum and maximum value, skewness, and kurtosis. The pattern of the distribution of the parameters was measured by skewness and kurtosis.

Table 1: Parameters with their transformation value

Parameters	Actual Values	Transform Into Category
Wind direction	in degree (°)	315° to 45° : North 45° to 135° : South 135° to 225° : East 225° to 315°: West
Rainfall amount	in mm	1mm – 10mm : (1) Light 11mm – 30mm : (2) Moderate >30mm : (3) Heavy

Trend

The Mann-Kendall (MK) trend test was used to assess the existence of a significant trend in the rainfall dataset. The Mann-Kendall test is part of statistical analysis that is suitable for all types of distributions, and outliers and missing values, as it is suitable for the rainfall data that is used in this study. This method is applied to find the trend of rainfall in Malaysia whether the trend is increasing or decreasing in time series (Güçlü, 2020). Mann-Kendall test statistic (S) can be obtained as written in Equation (1):

$$S=\sum_{i=2}^{n}\sum_{j=1}^{i-1} sign\left(x_i - x_j\right)$$

where n is the length of the time series x_i (i = 1, 2,…, n -1) and sign x_j ($j = i$ + 1, 2,…, n) are the data values and $sign\left(x_i - x_j\right)$ is the sign function as written in Equation 2.

$$sign\left(x_i - x_j\right) = \begin{cases} +1, \ if \ x_i - x_j > 0 \\ 0, \quad if \ x_i - x_j = 0 \\ -1, \ if \ x_i - x_j < 0 \end{cases} \tag{2}$$

Then, when the number of data points is greater than or equal to 10 ($n \geq 10$), the Mann-Kendall test is characterized by a normal distribution with mean value $E(S)$ = 0 and the variance value $Var(S)$ is equated as written in equation 3.

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{k=1}^{m} t_k\left(t_k - 1\right)\left(2t_k + 5\right)}{18} \tag{3}$$

where n is the number of data points, m is the number of tied groups and t_k denotes the number of ties in the kth tied groups. A tied group is a set of sample data having the same value. Then, the standardized Z statistic test can be derived using the equation as shown in equation 4:

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{Var(S)}}, & if \ S > 0 \\ 0, & if \ S = 0 \\ \dfrac{S+1}{\sqrt{Var(S)}} & if \ S < 0 \end{cases} \tag{4}$$

The result of the standardized Z statistic test is the main result in analysing the trend as it indicates an increasing or decreasing trend. For Z greater than 0, it results in an increasing trend. When Z is less than zero it indicates a decreasing trend. In addition, the Mann-Kendall test can also be generated by using R software.

The hypothesis of this test is:

The null hypothesis, H_0: *There is no trend in the data.*

The alternative hypothesis, H_1: *There is a trend in the data.*

Diurnal Variation

The diurnal variation of rainfall data was analysed using hourly rainfall data from 2018 to 2020. The distribution of diurnal cycles was presented to get the pattern of rainfall annually. Diurnal variation of rainfall was divided based on yearly seasonal rainfall that included the seasonal monsoon. The time-series data of rainfall was analysed based on average and maximum rainfall data to see the pattern of heavy rainfall in which month will be facing a high rainfall state. The categorization of the monsoon was based on the study done by Alias *et al.* (2016) :

(i) The Northeast monsoon: November until March

(ii) The Southwest monsoon: May until September

(iii) The two short inter-monsoons: April and October

The result of the diurnal analysis of rainfall in Malaysia helps the researcher to understand the pattern of rainfall annually.

Model Assessment

The main objective of this study is to apply discriminant analysis in the prediction of rainfall classification in selected locations in Malaysia. Discriminant analysis or also known as Fisher's Discriminant is a statistical method that is used to classify or group observations from the dependent variable into distinct groups of the quantitative independent variable. The objective of discriminant analysis is to predict group membership by a set of quantitative variables. If there are only two groups to be classified, the linear discriminant function is shown in Equation 5.

$$\hat{y} = a_1x_1 + a_2x_2 + \ldots + a_nx_n, \textit{ where } n = 1,2,\ldots \quad (5)$$

where $\hat{y}$ represents the discriminant function and a is a predictor.

In the case of more than two groups, a linear discriminant score is calculated. The linear discriminant score is a function of the population means for each of the groups as well as the pooled variance-covariance matrix. The function is provided in Equation 6.

$$\hat{d}_i(x) = \ln p_i + \bar{x}_i S^{-1}{}_{pooled} - \frac{1}{2}\bar{x}_i S^{-1}{}_{pooled} \bar{x}_i \quad (6)$$

where $\hat{d}_i$ represents the discriminant score and p is the probability.

The assumptions of data were checked beforehand to ensure the accuracy of the discriminant function obtained.

(i) Normality of the independent variables for each level in the group variables. The normality can be tested by using a cut-off point of 50% from the comparison of the Mahalanobis distance and chi-square or critical value.

(ii) Linearity of the dependent variables. The Pearson correlation coefficient was used to test the correlation between the variables for each location.

(iii) Multivariate homogeneity of variance between the groups. Levene's test was used to test the homogeneity of variance with the following testing:

The null hypothesis, H_0: $\sigma_1^2 = \sigma_2^2$ (Population variances for groups 1 and 2 are equal)

The alternative hypothesis, H_1: $\sigma_1^2 \neq \sigma_2^2$ (Population variances for groups 1 and 2 are not equal)

The null hypothesis is rejected when the significance values are less than 0.05. Since the assumption is having an equal variance, hence the aim is to not reject the null hypothesis. However, if the assumption is not met, it can be waived under some conditions. It is suggested that if the assumption is not met, which makes the data have unequal variance among the groups, the assumption can be waived if the size of the dataset is large (Cheung, 2018; Johnson & Wichern, 2014).

(iv) Multivariate homogeneity of covariance between groups. This assumption can be analysed using Box's M test. The compliance of this assumption can be waived if the size of the dataset is large (Cheung, 2018; Johnson & Wichern, 2014).

Performance Indicators

The performance of the classification function is assessed via its error rates (probabilities of misclassification). The Apparent Error

Rate (APER) is defined as the proportion of observations in the training set that are misclassified by the classification function. This method works better when the calculation uses the confusion matrix as shown in Table 2.

Table 2: Confusion matrix

		Predicted		
		π_1	π_2	
Actual	π_1	n_1c	$n_{1M} = n_1 - n_1c$	n_1
	π_2	$n_{2M} = n_2 - n_2c$	n_2c	n_2

Where π_1 and π_2 are the classifications:

n_1c is the number of π_1 items that are correctly classified as π_1 item
n_2c is the number of π_2 items that are correctly classified as π_2 item
n_{1M} is the number of π_1 items that are misclassified as π_2 item
n_{2M} is the number of π_2 items that are misclassified as π_1 item

The Apparent Error Rate (APER) is calculated as shown in Equation (7),

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{7}$$

which is recognized as the proportion of items in the training set that are misclassified (Johnson & Wichern, 2014). In general, the acceptable misclassification rate is 30% (Ahmat *et al*., 2019).

Model Validation

There are 2 steps to judge how good a model is, which are model verification and model validation. The model verification and validation processes ensure that a simulation model makes sense for specific research purposes. For this study, the model was verified using the APER to ensure the simulation model is complete and can be applied to the real problem. The proposed discriminant model for this study was further validated to determine how accurate the simulation model is as a representation of a real-world system for the simulation. The researcher calculates the cross-validation rate using this formula:

CV error rate: % of observations in the validation data which are misclassified by the classification function.

The function is considered valid when error rates are less than 30% (Ahmat *et al.,* 2019). The Discriminant Function Plot was also used to determine obvious distinctions between groups of data and whether is there any overlapping between the values.

Final Predictive Model

In this study, the rainfall was categorized into three based on previous research which were low, moderate and high intensity. New data can be substituted into the model which is subsequently assigned into which category it falls into. The categorization is based on the largest score that the predictive model obtained.

Results and Discussion

Descriptive Statistics

Table 3 summarizes the averages for all the meteorological parameters in Kuala Krai from 2018 to 2020. It can be observed that there was a difference in the hourly rainfall amount and the maximum daily rainfall amount. The average hourly rainfall amount was about the same throughout the year. However, the average maximum daily rainfall amount was slightly higher in 2020.

Table 4 lists the average of all the meteorological parameters in Kuala Terengganu from 2018 to 2020. The value of the means across all variables was almost equal throughout the years. Similar to Kuala Krai, the average hourly rainfall amounts were lower than the average maximum daily rainfall amount for the 3 years.

Table 5 summarizes the statistics for all the meteorological parameters in Temerloh. Similar to the 2 locations, the mean value of all variables was about the same for the 3 years.

Table 3: Summary statistics data in Kuala Krai

Measure	Year	Hourly Temperature (°C)	Hourly Humidity (%)	Hourly Wind Speed (m/s)	Hourly Rainfall Amount (mm)	Maximum Daily Rainfall Amount (mm)
Mean	2018	26.9	85.7	0.8	0.3	3.4
	2019	26.9	85.1	1.0	0.3	3.2
	2020	27.0	83.5	0.9	0.3	3.6
Minimum	2018	16.2	48.0	0.0	0.0	0.0
	2019	18.4	41.0	0.0	0.0	0.0
	2020	18.1	40.0	0.0	0.0	0.0
Maximum	2018	36.6	100.0	9.1	67.2	67.2
	2019	37.2	98.0	9.0	41.6	41.6
	2020	37.1	99.0	9.1	38.2	38.2

Table 4: Summary statistics of rainfall in Kuala Terengganu

Measure	Year	Hourly Temperature (°C)	Hourly Humidity (%)	Hourly Wind Speed (m/s)	Hourly Rainfall Amount (mm)	Maximum Daily Rainfall Amount (mm)
Mean	2018	27.6	82.06	1.89	0.32	3.72
	2019	27.9	81.11	2.12	0.25	3.16
	2020	27.9	82.36	2.02	0.34	3.56
Minimum	2018	19.6	50.0	0.0	0.0	0.0
	2019	21.4	43.0	0.0	0.0	0.0
	2020	22.0	53.0	0.0	0.0	0.0
Maximum	2018	34.4	100.0	10.6	52.6	52.6
	2019	34.4	100.0	9.3	44.0	44.0
	2020	34.5	100.0	7.8	84.0	84.0

Table 5: Summary statistics of rainfall data in Temerloh

Measure	Year	Hourly Temperature (°C)	Hourly Humidity (%)	Hourly Wind Speed (m/s)	Hourly Rainfall Amount (mm)	Maximum Daily Rainfall Amount (mm)
Mean	2018	27.4	82.3	0.8	0.2	3.1
	2019	27.7	81.8	0.9	0.2	2.9
	2020	27.6	82.6	0.8	0.2	3.5
Minimum	2018	20.5	37.0	0.0	0.0	0.0
	2019	20.4	36.0	0.0	0.0	0.0
	2020	21.2	43.0	0.0	0.0	0.0
Maximum	2018	36.6	100.0	6.2	63.2	63.2
	2019	36.9	100.0	5.1	48.6	48.6
	2020	36.6	100.0	6.1	63.6	63.6

Trend

The *p*-value of the Mann-Kendall test is based on the significance level of 95% and the *Z* value as the indication of an increasing or decreasing trend of rainfall. As depicted in Table 6, the *Z* statistics revealed that only two stations recorded an increasing trend throughout 2018-2020. The Kuala Krai station recorded a significant increasing trend every year, in contrast with the findings by Muhammad *et al*.(2021), where there was no trend found for monthly data from 2013 to 2019 in Kuala Krai ($p > 0.05$). Kuala Terengganu recorded a significant increasing trend only for the years 2018 and 2020. There was no significant trend in Temerloh throughout 2018-2020. Similarly, Chang *et al.* (2017) found no significant trend in rainfall for Temerloh for 40 years. According to him, the positive trends of rainfall in Pahang were indirectly or directly affected by Northeast Monsoon and Southwest Monsoon. Figure 3 displays the monthly trend of rainfall for Kuala Krai and Kuala Terengganu.

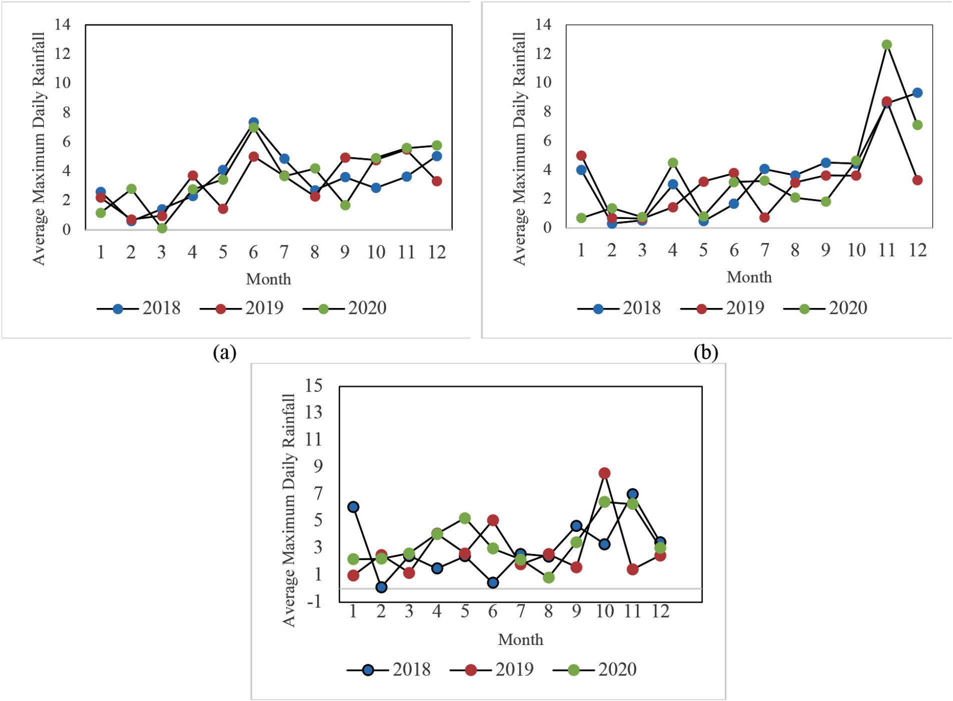
Figure 3 shows the trend of monthly rainfall in Kuala Krai and Kuala Terengganu from 2018 to 2020. Kuala Krai had an increasing trend and the wettest month was recorded in June 2018. The increasing trend could be observed during the Northeast monsoon starting in November, which is in line with a study by Muhammad *et al*. (2021) who concluded that seasonality affects the amount of rainfall in Kuala Krai where it is generally high in November and December each year. The significant increasing trend of rainfall in Kuala Terengganu started from February until December 2018 and 2020. The wettest month was recorded during the Northeast Monsoon, which was in November and December. The trend of rainfall in Temerloh fluctuated throughout the year, however, the trend was not significant from 2018 to 2020 compared to trends in Kuala Krai and Kuala Terengganu. This is comparable with the findings by Abu Bakar *et al.* (2020) and Mohd Ariff *et al.* (2019) who found that the differences are greatly influenced by the monsoon and intermonsoon seasons in Peninsular Malaysia.

Diurnal Variation

Figure 4 shows the graph of diurnal variation in Kuala Krai, Kuala Terengganu and Temerloh. The highest concentration of rainfall in Kuala Krai generally occurred during the late afternoon or early evening. Early in the morning, the pattern of rain variation was almost similar and started to increase during the afternoon. Throughout all monsoons, late inter-monsoon

Table 6: Mann-Kendall Trend test

Station	**2018**			
	***p*-value**	***Z***	**MK Trend**	**Test interpretation**
Kuala Krai	0.04674	1.9886	Increasing	There is a trend
Kuala Terengganu	0.003192	2.9486	Increasing	There is a trend
Temerloh	0.1148	1.5772	Increasing	No trend
Station	**2019**			
	***p*-value**	***Z***	**MK Trend**	**Test Interpretation**
Kuala Krai	0.03352	2.1257	Increasing	There is a trend
Kuala Terengganu	0.1926	1.3029	Increasing	No trend
Temerloh	0.6312	0.48001	Increasing	No trend
Station	**2020**			
	***p*-value**	***Z***	**MK Trend**	**Test Interpretation**
Kuala Krai	0.01117	2.5372	Increasing	There is a trend
Kuala Terengganu	0.007488	2.6743	Increasing	There is a trend
Temerloh	0.1926	1.3029	Increasing	No trend



(a) (b)

(c)

Figure 3: Trend of monthly rainfall in (a) Kuala Krai, (b) Kuala Terengganu and (c) Temerloh

during October was found to have the maximum concentration of rainfall. The maximum rainfall normally took place in the evening during the Northeast monsoon. During other monsoon seasons, rainfall mostly occurred in the late afternoon and early evening.

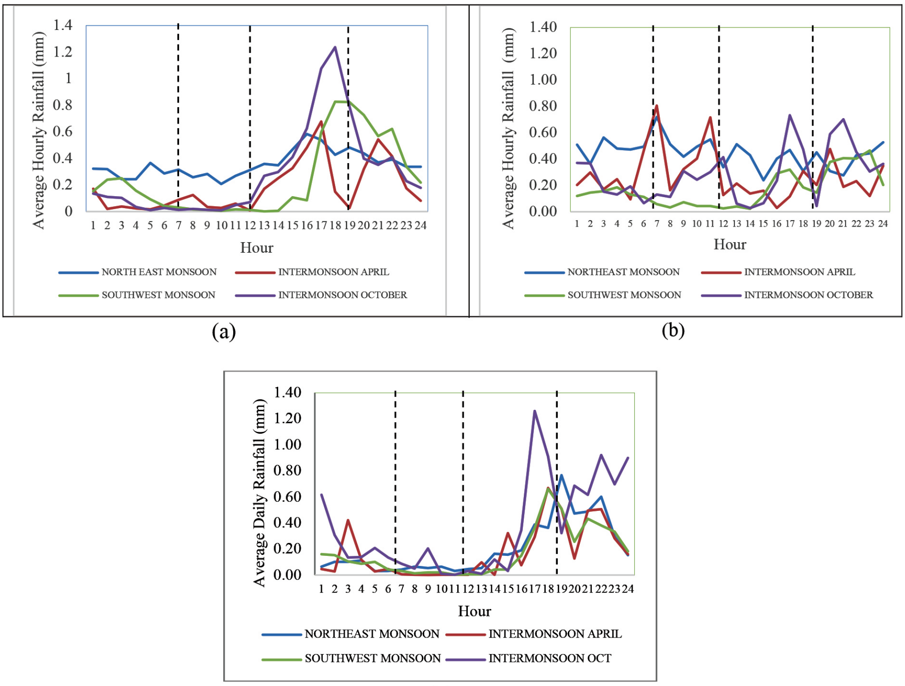
In Kuala Terengganu, the concentration of rainfall normally occurred throughout the day. During the inter-monsoon in April, the highest concentration of rainfall occurred in the morning while for inter-monsoon in October, the high concentration of rainfall struck in the late evening. Kuala Terengganu received the least amount of rainfall early in the morning during the Southwest monsoon but recorded a certain amount of rainfall throughout the entire day for other monsoon seasons.

In Temerloh, the pattern of rainfall trended to late in the afternoon and early evening for all monsoon seasons. The least amount of rainfall was recorded from early morning until the afternoon with the average hourly rainfall below 0.2 mm. During the late inter-monsoon season in October, the highest rainfall concentration was in the late afternoon. During the Northeast monsoon, most of the rainfall occurred during the late afternoon and early evening. The pattern of rainfall is roughly the same throughout the monsoons except during intermonsoon in October.

Classification of Rainfall

Table 7 provides the value of Wilk's Lambda and the significance values for all three locations. The null hypothesis would be no significant relationship between independent variables and dependent variables vs the alternative hypothesises that the relationships are significant The significant p = 0.000 less than 0.05 for all the locations, hence, the relationships among the variables were deemed significant.

Table 8 tabulates the discriminant equations for Kuala Krai, Kuala Terengganu and Temerloh. The wind speed and humidity were



(a)

(b)

(c)

Figure 4: Diurnal variation for (a) Kuala Krai, (b) Kuala Terengganu and (c) Temerloh

identified as the most significant contributors affecting rainfall in Kuala Krai and Temerloh. On the other hand, for Kuala Terengganu, the most significant factors were wind direction and temperature. To compare, Sulaiman *et al*. (2020) used Principal Component analysis for rainfall data in Kelantan and Terengganu and also found that component 1 consisted of temperature, precipitation, humidity and solar and component 2 comprised of wind and temperature. This was almost similar to the significant contributors affecting the level of rainfall in Kuala Krai and Kuala Terengganu.

After the discriminant equations have been identified, the classification of rainfall intensity of either light, moderate or heavy can be done via classification scores. The intensity of rainfall will be classified into the group for which it has the highest classification score.

Performance Indicator

Two error rates were used namely, Apparent Error Rate (APER) and Cross-Validation Rate. APER was used to check for misclassification in the model obtained while Cross Validation Rate was used to check for misclassification in the validation dataset. As shown in Table 9, the discriminant functions were considered good since all the misclassification rates were well below 24%. In general, the acceptable misclassification rate is about 30%.

Table 7: Value of Wilk’s Lambda based on variables in a dataset of Kuala Krai, Kuala Terengganu and Temerloh

Locations	Variable	F	Sig. value	Conclusion
Kuala Krai	Rainfall Classification	37.684	0.000	Significant
Kuala Terengganu	Rainfall Classification	59.560	0.000	Significant
Temerloh	Rainfall Classification	92.822	0.000	Significant

Table 8: The discriminant equations for Kuala Krai, Kuala Terengganu and Temerloh

Location	Discriminant Equations
Kuala Krai	$\widehat{d}_{light} = -0.001ZTemperature - 0.005ZHumidity - 0.004WindDirection - 0.11ZWindSpeed - 1.099$ $\widehat{d}_{moderate} = -0.008ZTemperature + 1.845ZHumidity + 0.671WindDirection + 1.304ZWindSpeed - 2.151$ $\widehat{d}_{heavy} = -0.956ZTemperature + 1.062ZHumidity + 0.742WindDirection + 1.476ZWindSpeed - 2.326$
Kuala Terengganu	$\widehat{d}_{light} = -0.007ZTemperature + 0.006ZHumidity - 0.016WindDirection - 0.006ZWindSpeed - 1.099$ $\widehat{d}_{moderate} = 0.951ZTemperature - 0.419ZHumidity + 2.076WindDirection + 0.067ZWindSpeed - 2.397$ $\widehat{d}_{heavy} = 1.239ZTemperature - 0.407ZHumidity + 2.151WindDirection - 0.146ZWindSpeed - 2.822$
Temerloh	$\widehat{d}_{light} = 0.029ZTemperature + 0.019ZHumidity - 0.006WindDirection - 0.006ZWindSpeed - 1.099$ $\widehat{d}_{moderate} = 0.066ZTemperature + 3.055ZHumidity + 0.631WindDirection + 2.668ZWindSpeed - 3.573$ $\widehat{d}_{heavy} = -1.073ZTemperature + 2.332ZHumidity + 0.967WindDirection + 2.940ZWindSpeed - 4.419$

Table 9: Apparent Error Rate for Kuala Krai, Kuala Terengganu and Temerloh

Locations	Apparent Error Rate (<30%)	Cross Validation Rate
Kuala Krai	23.6	21.28
Kuala Terengganu	23.2	23.00
Temerloh	13.3	12.97

Simulation

For illustration, Table 10 shows the calculation of the discriminant score using discriminant equations obtained. The illustration data used was for 24 November in Kuala Krai. One of the extraordinary flood events recorded on 24 November involved several locations in Kuala Krai, with the highest maximum rainfall recorded in Kelantan. One of the causes of floods in Kelantan was heavy rainfall coupled with other factors (Jabatan Pengairan dan Saliran, 2021). The results in Table 8 show a good agreement with the actual data.

Table 10: Simulation using Discriminant Equations for Kuala Krai

Date	Rainfall	ZTemp	ZHumid	Wind Direct	ZWind Speed	D_{light}	$D_{moderate}$	D_{heavy}	Rainfall Initial Category	Rainfall Predicted Category
24/11	18.20	-0.38	0.90	1	1.64	-1.29	2.32	2.15	Moderate	Moderate

Conclusion

The research identified that the main contributing factors for rainfall intensity in Kuala Krai and Temerloh were wind speed and humidity, while in Kuala Terengganu they were wind direction and temperature. Kuala Krai and Temerloh are in the centre of Kelantan and Pahang respectively. Kuala Terengganu, on the other hand, is near the beach thus, it explains why the temperature and the wind direction heavily affect the intensity of rainfall in Kuala Terengganu and not in Kuala Krai and Temerloh.

The misclassification rate shows that the discriminant functions obtained were good since both the misclassification rates were less than 30%. The simulation results show a good agreement with real conditions that occurred in Kuala Krai on 24 November 2020, as reported in Laporan Banjir Tahunan 2020 (Jabatan Pengairan dan Saliran, 2021). Therefore, the discriminant functions can be used to classify the level of rainfall intensity in Kuala Krai, Kuala Terengganu and Temerloh. The findings of this study will be beneficial to future researchers, such as meteorologists and climate scientists, who might examine other factors that can determine more natural disasters, such as floods and drought.

Acknowledgements

The authors would like to thank the Malaysian Meteorological Department for providing the meteorological data in this study.

References

Abu Bakar, M. A., Ariff, N. M., Jemain, A. A., & Nadzir, M. S. M. (2020). Cluster analysis of hourly rainfalls using storm indices in Peninsular Malaysia. *Journal of Hydrologic Engineering*, *25*(7), 1-11. https://doi.org/10.1061/(asce)he.1943-5584.0001942

Ahmad, F., Ushiyama, T., & Sayama, T. (2017). *Determination of Z-R Relationship and Inundation Analysis for Kuantan River*. Malaysian Meteorological Department (MMD). https://www.met.gov.my/data/research/researchpapers/2017/researchpaper_201702.pdf

Ahmat, H., Musa, Nor Syahida, Nazamid, N., & Zaharin, Nursyahirah Amirah. (2019). Classification of high and low level of Pm 10 concentrations in Klang and Shah Alam ,. *Malaysian Journal of Computing*, *4*(2), 325-334.

Alias, N. E., Mohamad, H., Chin, W. Y., & Yusop, Z. (2016). Rainfall analysis of the Kelantan big yellow flood 2014. *Jurnal Teknologi*, *78*(9-4), 83-90. https://doi.org/10.11113/JT.V78.9701

Arief Rosyidie. (2013). Banjir: Fakta dan dampaknya, serta pengaruh dari perubahan guna lahan. *Journal of Regional and City Planning*, *24*(3), 241-249. https://doi.org/10.5614/JPWK.2013.24.3.1

Aziz, A., Harun, N. A., Makhtar, M., Syed, F., Abdullah, Jusoh, J. A., & Zakaria, Z. A. (2016). A conceptual framework for predicting flood area in Terengganu during monsoon season using association rules. *Journal of Theoretical and Applied Information Technology*, *87*(3), 512-519.

Brereton, R. G. (2015). The Mahalanobis distance and its relationship to principal component scores. *Journal of Chemometrics*, *29*(3), 143-145. https://doi.org/10.1002/cem.2692

Chang, C. K., Ghani, A. A., & Othman, M. A. (2017). Homogeneity testing and trends analysis in long term rainfall data for Sungai Pahang river basin over 40 years records. *Proceedings of the 37th IAHR World Congress*, *August 13-18*, 4197-4203.

Cheung, M. W. L. (2018). Computing multivariate effect sizes and their sampling covariance matrices with structural equation modelling: Theory, examples, and computer simulations. *Frontiers in Psychology*, *9*(Aug), 1387.

Department of Irrigation and Drainage, M. of E., & W. (n.d.). *Rainfall data - The official web of public infobanjir*. Retrieved September 12, 2022, from https://publicinfobanjir.water.gov.my/hujan/data-hujan/?lang=en

Elfithri, R., Halimshah, S., Pauzi Abdullah, M., Mokhtar, M., Ekhwan Toriman, M., Fuad Embi, A., Abdullah, M., Yook Heng, L., Nizam Ahmad Maulud, K., Salleh, S., Maizan, M., & Mohamad Ramzan, N. (2017). Pahang flood disaster : The potential flood drivers. *Malaysian Journal Geosciences (MJG)*, *1*(1), 34-37.

Güçlü, Y. S. (2020). Improved visualization for trend analysis by comparing with classical Mann-Kendall test and ITA. *Journal of Hydrology*, *584*, 124674.

Hashim, M., Nayan, N., Setyowati, D. L., Said, Z., Hanifah, Mahat, & Saleh, Y. (2021). Analysis of water quality trends using the mann-kendall test and sen's estimator of slope in a Tropical River Basin. *Pollution*, *7*(4), 933-942.

Ismail, W. R., & Haghroosta, T. (2018). Ismail and Haghroosta 2018. *Research in Marine Sciences*, *3*(1), 231-244.

Jabatan Pengairan dan Saliran. (2021). Laporan banjir tahunan. In *Bahagian Sumber Air & Hidrologi*.

Johnson, R. A., & Wichern, D. W. (2014). Applied multivariate statistical analysis. In *The Mathematical Gazette* (6th ed., Vol. 72, Issue 462). Pearson Education Limited. https://doi.org/10.2307/3619964

Keskin, A. I., Dincer, B., & Dincer, C. (2020). Exploring the impact of sustainability on corporate financial performance using discriminant analysis. *Sustainability*, *12*, 2346. https://doi.org/10.3390/SU12062346

Khan, N., Pour, S. H., Shahid, S., Ismail, T., Ahmed, K., Chung, E. S., Nawaz, N., & Wang, X. (2019). Spatial distribution of secular trends in rainfall indices of Peninsular Malaysia in the presence of long-term persistence. *Meteorological Applications*, *26*(4), 655-670.

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, *32*(1), 1-10. https://doi.org/10.5334/irsp.289

Maisarah, W., Ibadullah, W., Tangang, F., Juneng, L., & Fairudz Jamaluddin, A. (2019). Practical predictability of the 17 December 2014 heavy rainfall event over East Coast of Peninsular Malaysia using WRF model (Kebolehramalan Praktikal Peristiwa Hujan Lebat pada. *Sains Malaysiana*, *48*(11), 2297-2306.

Mayowa, O. O., Pour, S. H., Shahid, S., Mohsenipour, M., Harun, S. Bin, Heryansyah, A., & Ismail, T. (2015). Trends in rainfall and rainfall-related extremes in the East Coast of Peninsular Malaysia. *Journal of Earth System Science*, *124*(8), 1609-1622.

Mignot, E., Li, X., Dewals, B. J., & Dewals, B. (2019). Experimental modelling of urban flooding: A review. *Journal of Hydrology*, *568*. https://doi.org/10.1016/j.jhydrol.2018.11.001ï

Mohd Ariff, N., Aftar Abu Bakar, M., Faridah Syed Mahbar, S., Shahrul Mohd Nadzir, M., Operasi Cuaca, P., Nasional, G., Meteorologi Malaysia, J., & Tenaga, K.

(2019). Clustering of rainfall distribution patterns in Peninsular Malaysia using time series clustering method. *Malaysian Journal of Science*, *38*(Sp2), 84-99. https://doi.org/10.22452/MJS.SP2019NO2.8

Muhammad, M., Ibrahim, Q. A. Q., Ghani, M. S. M., Jemali, N., & Awang, N. R. (2021). Spatio-temporal analysis of rainfall data in Kuala Krai Kelantan. *IOP Conference Series: Earth and Environmental Science*, *842*, 012023. https://doi.org/10.1088/1755-1315/842/1/012023

Muhammad, M. K. I., Nashwan, M. S., Shahid, S., Ismail, T. bin, Song, Y. H., & Chung, E. S. (2019). Evaluation of empirical reference evapotranspiration models using compromise programming: A case study of Peninsular Malaysia. *Sustainability*, *11*(16), 655-670.

Nor, S. M. C. M., Shaharudin, S. M., Ismail, S., Zainuddin, N. H., & Tan, M. L. (2020). A comparative study of different imputation methods for daily rainfall data in East-Coast Peninsular Malaysia. *Bulletin of Electrical Engineering and Informatics*, *9*(2), 635-643. https://doi.org/10.11591/eei.v9i2.2090

Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N., & El-Shafie, A. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal*, *12*(2), 1651-1663. https://doi.org/10.1016/J.ASEJ.2020.09.011

Shaharudin, S., & Ahmad, N. (2019). Classification of daily torrential rainfall patterns based on a robust correlation measure. *International Journal of Advanced Science and Technology*, *28*(8s), 251-259. http://sersc.org/journals/index.php/IJAST/article/view/878

Sulaiman, N. A., Milleana Shaharudin, S., Hila Zainuddin, N., & Aimi Mohd Najib, S. (2020). Improving support vector machine rainfall classification accuracy based on kernel parameters optimization for statistical downscaling approach. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(1.4), 652-657. https://doi.org/10.30534/ijatcse/2020/9191.42020

Tam, T. H., Abd Rahman, M. Z., Harun, S., Hanapi, M. N., & Kaoje, I. U. (2019). Application of satellite rainfall products for flood inundation modelling in Kelantan River Basin, Malaysia. *Hydrology, 6*(4), 95. https://doi.org/10.3390/HYDROLOGY 6040095

Wong, C. L., Venneker, R., Uhlenbrook, S., Jamil, A. B. M., & Zhou, Y. (2009). Variability of rainfall in Peninsular Malaysia. *Hydrology and Earth System Sciences Discussions*. 6, 5471-5503, https://doi.org/10.5194/hessd-6-5471-2009, 2009

Wong, Chee Loong, Liew, J., Yusop, Z., Ismail, T., Venneker, R., & Uhlenbrook, S. (2016). Rainfall characteristics and regionalization in Peninsular Malaysia based on a high-resolution gridded data set. *Water*, *8*, 500.

Ziarh, G. F., Shahid, S., Ismail, T. Bin, Asaduzzaman, M., & Dewan, A. (2021). Correcting bias of satellite rainfall data using a physical empirical model. *Atmospheric Research*, *251*, 105430.