Ain Shams Engineering Journal 12 (2021) 1651–1663

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Electrical Engineering

Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia

Wanie M. Ridwan [a,b], Michelle Sapitang [a,b], Awatif Aziz [a], Khairul Faizal Kushiar [c], Ali Najah Ahmed [d,*], Ahmed El-Shafie [e,f]

[a] UNITEN R&D Sdn. Bhd., Universiti Tenaga Nasional (UNITEN), Kajang 43000, Selangor Darul Ehsan, Malaysia
[b] Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor, Malaysia
[c] Asset Management Department, Generation Division, Tenaga Nasional Berhad, 59200 Kuala Lumpur, Malaysia
[d] Institute for Energy Infrastructure (IEI), Universiti Tenaga Nasional (UNITEN), Kajang 43000, Selangor Darul Ehsan, Malaysia
[e] Department of Civil Engineering, Faculty of Engineering, University of Malaya (UM), 50603 Kuala Lumpur, Malaysia
[f] National Water Center (NWC), United Arab Emirates University, Al Ain P.O. Box. 15551, UAE

A R T I C L E I N F O

Article history:
Received 1 July 2020
Revised 7 September 2020
Accepted 9 September 2020
Available online 7 November 2020

Keywords:
Forecasting rainfall
Machine learning algorithms
Boosted decision tree regression
Decision forest regression
Neural network regression
Bayesian linear regression

A B S T R A C T

Rainfall plays a main role in managing the water level in the reservoir. The unpredictable amount of rainfall due to the climate change can cause either overflow or dry in the reservoir. In this study, several models and methods were applied to predict the rainfall data in Tasik Kenyir, Terengganu. The comparative study was conducted focusing on developing and comparing several Machine Learning (ML) models, evaluating different scenarios and time horizon, and forecasting rainfall using two types of methods. Data involved for this research consist of taking the average rainfall from 10 stations around the study area using Thiessen polygon to weight the station area and projected rainfall. The forecasting model uses four different ML algorithms, which are Bayesian Linear Regression (BLR), Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR) and Neural Network Regression (NNR). On the other hand, the rainfall was predicted on different time horizon by using different ML's algorithms which is method 1 (M1): Forecasting Rainfall Using Autocorrelation Function (ACF) and method 2 (M2): Forecasting Rainfall Using Projected Error. In M1, the best regression developed for ACF is BDTR since it has the highest coefficient of determination, R2, after tuning the hyperparameter. The results show coefficient between 0.5 and 0.9 with the highest of each scenarios for daily (0.9739693), weekly (0.989461), 10-days (0.9894429) and monthly (0.9998085). In M2, overall model performances show that normalization using LogNormal is preferably giving a good result of each categories except for 10-days with BDTR and DFR are the most acceptable result than NNR and BLR. It is concluded that, two different methods have been applied with different scenarios and different time horizons, and M1 shows a rather high accuracy than M2 using BDTR modeling.

1. Introduction

1.1. Background

Malaysia is described as a country that has hot climate all through of the year since it is placed close to the equator [1]. Climate change causes parts of the water cycle accelerate as global warming temperatures raise the rate of evaporation around the world. Evaporation, also called evapotranspiration, is defined as the combination of evaporation and plant transpiration from the land or ground to the atmosphere. The factors that affect the evapotranspiration are air temperature, wind speed, vapor pressure,

* Corresponding author.
E-mail addresses: ridwaniyah.ridwan@uniten.edu.my (W.M. Ridwan), michelle.sapitang@uniten.edu.my (M. Sapitang), awatif.aziz@uniten.edu.my (A. Aziz), mahfoodh@uniten.edu.my (A.N. Ahmed), elshafie@um.edu.my (A. El-Shafie).
Peer review under responsibility of Ain Shams University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.asej.2020.09.011
2090-4479/

Notation

M1	Method 1
M2	Method 2
ML	Machine Learning
ECR	East Coast Region
GIS	Geographical Information System
PCA	Principal Component Analysis
RF	Random Forest
BLR	Bayesian Linear Regression
BDTR	Boosted Decision Tree Regression
DFR	Decision Forest Regression
NNR	Neural Network Regression
MLP	Multiple Layer Perceptron
SMFM	Single Mode Forecasting Model
MMFM	Multiple Mode Forecasting Model
ANFIS	Neuron Fuzzy Inference System
SFLA	Shuffled Frog-Leaping Algorithm
SVR	Support Vector Regression
SVR-KHA	Support Vector Regression with Hybrid of Krill-Herd Algorithm
ENN-KHA	Elman Neural Network with Hybrid of Krill-Herd Algorithm
ENN	Elman Neural Network
SVR-FA	Support Vector Regression with Hybrid of Firefly Algorithm
ENN-FA	Elman Neural Network with Hybrid of Firefly Algorithm
MLP-ANN	Multilayer Perceptron Neural Network
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
RSE	Relative Squared Error
RAE	Relative Absolute Error
R	Coefficient of Determination / Correlation Coefficient
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
mm	millimeter
ha	hectare
m	meter

relative humidity, soil moisture, type of crop and crop growth season. On average, more evaporation leads to high precipitation rates and the impacts can be seen in many parts of Malaysia, but it is not evenly distributed. Some areas may experience more massive than average precipitation and the other regions may become prone to droughts as the traditional locations of monsoon occurs are in the eastern part. The weather and climate are commonly described through a few meteorological parameters and the foremost imperative meteorological parameter is rainfall intensity or rainfall amount [2]. Exuberant rain usually causing surges and avalanches, also leads to natural disaster. One of the major focuses of climate change study is to understand whether there is a change in the occurrence frequency and strength of heavy rainfall events. Moreover, rainfall form main input to the river basin where it affected the water capacity and release of a stream particularly during the torrential rainfall event.

The rainfall-runoff relationship is one of the most complex hydrological phenomena due to existence of complex non-linear connection within the change of precipitation into runoff. This action is very challenging to assimilate, owing to the existence of extensive number of factors included in the trial of physical process [4]. Subsequently, its definite modelling is imperative for water resource improvement and management and forecast of climate change like dry season and surges [5]. Based on the involvement of physical angles, rainfall—runoff models are classified as either system theoretic model or physical-based models. The physical -based models require the significant data almost the system mechanism as well as its parameters but different with system theoretical models which do not concern much about the physical processes of the problem because these models based on the rainfall and runoff data. Its look for characterize nonlinearity and non-stationary conduct from those information by the utilize of transmission functions [6]. Based on [7], Artificial Neural Network (ANN) have received global attention for rainfall-runoff modelling because of their capability to capture high degree of non-linearity and climate change of relationship between the hydrological variables without fully understanding the processes beneath. Meanwhile, this study makes a difference to constructed and compared four new different ML techniques which is Neural Network Regression (NNR), Boosted Decision Tree Regression (BDTR), Bayesian Linear Regression (BLR), and Decision Forest Regression (DFR) in forecasting rainfall.

The monsoon occurs at the East Coast of Peninsular Malaysia, which is Kelantan, Terengganu and Pahang, but this study focus on the surrounding area of Tasik Kenyir, Hulu Terengganu. In conjunction with describing the weather and climate of Tasik Kenyir, the nearest meteorological station was used to acquire data. The parameters considered were rainfall amount in millimeters (mm) with daily data covers the period from year 1985 to year 2019. Late December 2014, Terengganu experienced its most extreme rainfall events which attributed to climate change [8]. Therefore, the present of this study is to predict the amount of rainfall for future to prevent and help in flood forecasting.

The East Coast Region (ECR) experienced a few monsoon seasons which known as southwest monsoon (dry seasons), northeast monsoon (wet seasons) and inter monsoon [9]. The southwest monsoon occurs from May to August, northeast monsoon occurs from November to February and inter monsoon occurs from September to October and March to April [10]. Climate of Terengganu is best described by different monsoon seasons. The data management was performed in dividing the data into five characteristic which is daily, weekly, 10-days, and monthly.

1.2. Literature review

In forecasting rainfall, there are two dominant approaches which is conceptual modelling and system theoretical modelling [11,12]. Conceptual modelling is widely applied in hydrological forecasting because it considered to estimate within the physical mechanism which oversee the hydrologic process and ordinarily based on the features and expertise of a catchment. Unfortunately, this approach may not be attainable for rainfall forecasting due to critical calibration data to reasoning rainfall is not effectively to be collected and the work of rainfall calculations need advanced numerical tool [13]. System theoretical methods are applied mapping models to define the relationship between inputs and outputs while neglecting the physical structure processes. The most popular approaches is ARMAX by [14] in time series forecasting however this model incapable to predict changes that are not based on the past data, especially for the nonlinearity of rainfall correlated variables.

Recently, because of the numerous progress within the fields of pattern recognition methodology, there is various tools to forecast rainfall easily rather than previously used conventional approach of linear mathematical relationships supported operator experience, mathematical curves and guidelines [15]. Machine Learning tool is widely applied to unravel hydrological problems including rainfall forecasting. The important of this modelling is that the skill of the software to plot the input-output patterns without aforementioned expertise of the factors affecting the forecast parameters [16,17]. Subsequently, researchers have since started to hone this ML to predict form of the modelling approaches and parameter to improve precision and unwavering quality of depict the predicting model. In addition, researcher have used ML for various parameter such as predicting daily streamflow using multi-layer perceptron (MLP) [18] and neuro-fuzzy inference system (ANFIS) with shuffled frog-leaping algorithms (SFLA) [19], predicting daily evapotranspiration using support vector regression (SVR) [20], estimating of solar radiation using support vector regression (SVR) with hybrid of krill-herd algorithm (SVR-KHA) [21] and estimating soil temperature using support vector regression (SVR) and elman neural network (ENN) with hybrid of firefly algorithm (SVR-FA, ENN-FA) and krill herd algorithm (SVR-KHA, ENN-KHA) [22] and predicting sea-level using multilayer perceptron neural network (MLP-ANN) and ANFIS [23].

Researcher [24–26] are used ANN to forecast rainfall with each different method where [24] construct rainfall simulation model and provide accurate rainfall forecasting information, temporal and spatial distribution, [25] used short-term rainfall for urban catchment and the result show that the ANN model with lower lag outflanked in terms of forecasting exact index, [26] forecast daily rainfall with resilient propagation learning algorithm and the numerical results shown that proposed model is predominant to a numerous regression model in terms of forecasting precision indices. [27] forecast rainfall using seasonal ARIMA by arrange to monthly rainfall data and it turns out ARIMA (0,0,1)(1,1,1) was to be the most effective to predict future precipitation with a 95% confidence interval. [28] aim to compare random forest (RF) and support vector machine (SVM) for real time radar derived rainfall forecasting with two different forecast models which is single mode forecasting model (SMFM) and multiple mode forecasting model (MMFM). The result show that SVN based SMFM is more effective than RF based SMFM because in most cases RF based SMFM underestimates the observed radar derived rainfalls.

Meanwhile, researcher [29] use Emotional Neural Network (ENN) and Artificial Neural Network (ANN) for modelling rainfall-runoff in the Sone Command, Bihar as this area experiences flood due to heavy rainfall. Then ENN model give the best result for rainfall-runoff discharge with determination of correlation, R^2 = 0.879. Researcher [30] conclude that Bayesian Network (BN) and ANN models can be successfully water quality modelling and forecasting but the most great in dealing with time series is BN including incomplete or missing data. Therefore, this study constructed and compared four new different ML techniques which is Boosted Decision Tree Regression (BDTR), Bayesian Linear Regression (BLR), Neural Network Regression (NNR) and Decision Forest Regression (DFR) in forecasting rainfall.

1.3. Problem statement

Rainfall forms the primary input to the river basin, affecting the water capacity a stream, particularly during the torrential rainfall event [3]. Moreover, one of the major focuses of climate change study is to understand whether there is an extreme changes in the occurrence and frequency of heavy rainfall events. The accuracy level of the ML models used in predicting rainfall based on historical data has been one of the most critical concerns in hydrological studies [31]. An accurate ML forecasting model could give early alerts of severe weather to help prevent natural disasters and destruction. Hence, there is needs to develop ML algorithms capable in predicting rainfall with acceptable level of precision and in reducing the error in the dataset of the projected rainfall from climate change model with the expected observable rainfall.

1.4. Objectives

The major aim of the study is to develop and compare several ML models, namely Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR), Bayesian Linear Regression (BLR) and Neural Network Regression (NNR) for the prediction of rainfall in Kenyir Lake, Hulu Terengganu. Secondly, this study also assesses and evaluate different scenarios and time horizons to find the most accurate and reliable model and the effectiveness of these algorithms in learning the sole input of rainfall pattern. Thirdly, the study also uses two method to predict the rainfall which are forecasting rainfall using Autocorrelation Function (ACF) and forecasting rainfall using projected error. This study is a vital to enhance the water management, especially in Kenyir Lake area.

2. Materials and methods

2.1. Case study and data description

Data for this study consist of 10 stations covering the Kenyir Lake area in Hulu Terengganu, Terengganu as shown in Fig. 1.

The basic statistical parameters are presented in Table 1. Table 1 shows that the standard deviation for each station varies between 18 mm and 30 mm. All the stations have recorded rainfall of 0 mm as the minimum and the maximum rainfall is 539.5 mm in Station 7, followed by Station 1 (455.5 mm) and Station 2 (440 mm). The maximum rainfall range for all the station in between the range of 325.5 mm to 539.5 mm.

For the 10 stations, each station should have 3,455 daily rainfall values. But most of the rainfall station values (except for two stations; Station 1 and Station 7) are not indicated due to sensor or technical issue. The most common approach is to delete individuals and/or variables containing missing observations. However, this loss of information reduces the ability to detect patterns and introduces biases if data are not missing completely at random [32]. Principal Component Analysis commonly simplifies as PCA, is the oldest and widely used machine learning technique. It explores correlations between variables and determines the combination of values that better represents variations in outcomes. Such integrated feature values are used to create a more compact feature space called the principal (key) components [33,34].

In this study, the Thiessen Polygon method calculates the proximity area around the point with other points. It also can help to calculate station rainfall weight and average rainfall based on each station. However, this method is suitably used for flat areas [23]. The total area for each rainfall stations using the Thiessen Polygon method is shown in Table 2 below.

$$P = \frac{P_1A_1 + P_2A_2 + P_3A_3 \cdots . + P_nA_n}{A_1 + A_2 + A_3 \cdots + A_n} \quad (1)$$

where P_1 is the daily rainfall values for first station, A_1 is the area of Thiessen polygon area for first station, P_n is the daily rainfall values at nth station and A_n = the area of Thiessen polygon area at nth station.

The average rainfall for the 10 stations is then calculated using the Equation in (1) referred to Lal [35] to get the average of daily rainfall. The basic statistical parameters for 3,455 daily value of average rainfall from January 2010 until June 2019, namely, mean, median, standard deviation (S.D.), range, minimum and maximum of the average rainfall data, are presented in Table 3. For this study, the data also use projected rainfall to predict the error. From

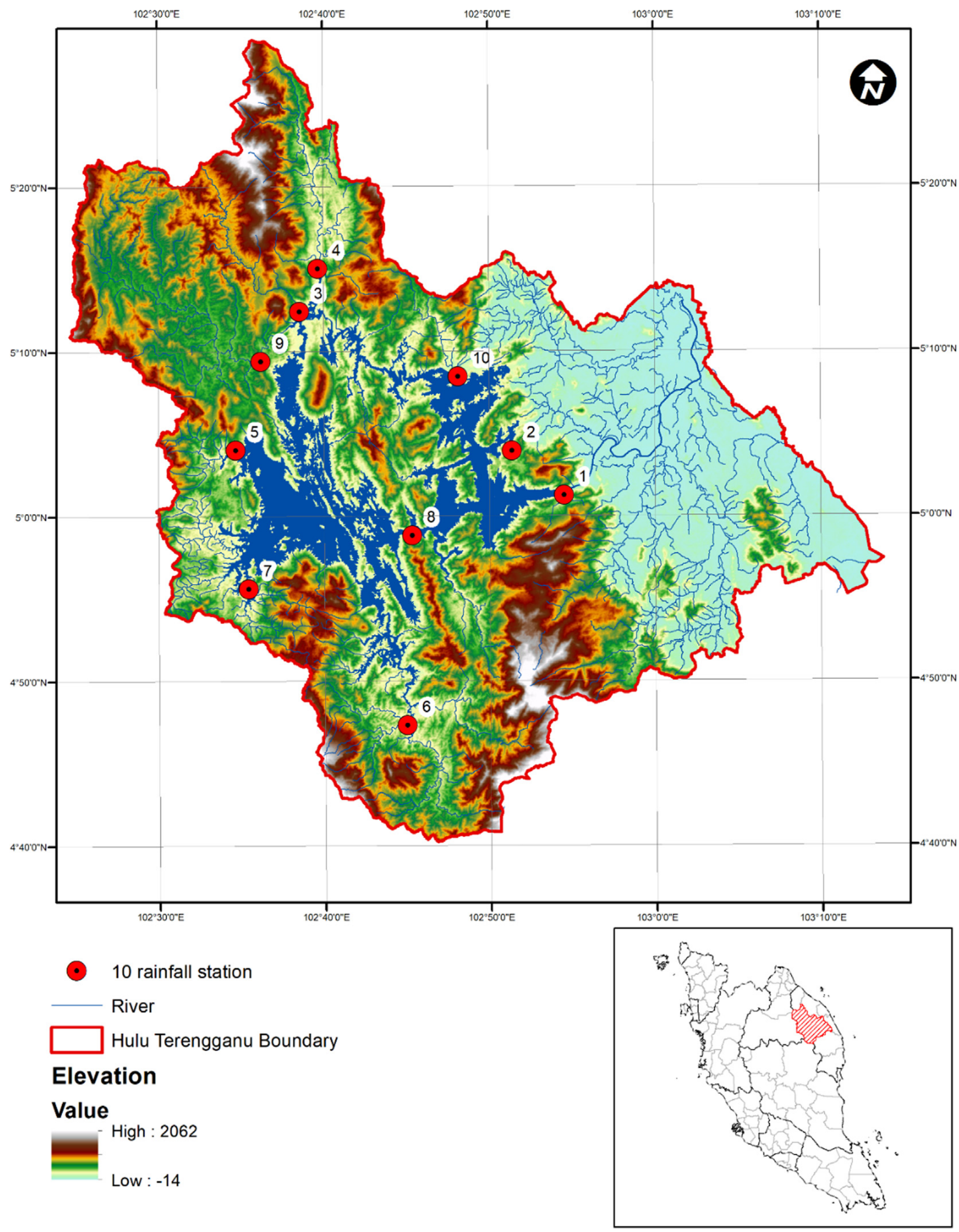

Fig. 1. Rainfall points within Kenyir Lake areas.

Table 1
Rainfall station with its'descriptive analysis.

Station No.	Mean (mm)	Standard Deviation (mm)	Range (mm)	Minimum (mm)	Maximum (mm)
1	12.73864	30.26158	455.5	0	455.5
2	11.26588	28.4316	440	0	440
3	9.788515	23.91329	344.5	0	344.5
4	10.8276	26.01967	420	0	420
5	8.882498	22.47946	325.5	0	325.5
6	7.987888	18.84153	352.5	0	352.5
7	9.811867	26.31595	539.5	0	539.5
8	9.587695	23.36292	360	0	360
9	9.299299	21.94225	339.5	0	339.5
10	10.92149	28.89012	424	0	424

Table 2
Polygon station area of for the Kenyir Lake catchment.

Station No.	Polygon Area (km^2)
1	228.2137
2	172.3874
3	129.5302
4	153.1546
5	173.3922
6	413.7608
7	211.0905
8	354.8176
9	150.7165
10	331.7657
	$\sum$ Area =2318.8292

Table 3, there is a huge difference for the maximum rainfall, about 517 mm.

2.2. Models used for forecasting

The forecasting model uses four different ML algorithms, which are Bayesian Linear Regression (BLR), Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR) and Neural Network Regression (NNR).

A BDTR is a classic method to create an ensemble of regression trees where each tree is dependent on prior tree [36]. In a simple word, it is an ensemble learning method during which the errors of the primary tree will be corrected by the second tree, the third tree corrects for the errors of the primary and second trees, then so on. Predictions are based together on the whole set of trees which makes the prediction. The BDTR shows an exceptionally great in dealing with tabular data [37]. The advantages of BDTR is it robust to missing data and normally allocates feature significance scores. Usually BDTR perform better than DFR because it appears to be the chosen method with slightly better performance than DFR in Kaggle competition [38]. Unlike DFR, BDTR may be more prone to overfitting because the main purpose is to reduce bias and not variance. BDTR takes a longer time to develop model since it has many hyperparameters to tune and trees are built sequentially [39].

A DFR is an ensemble of randomly trained decision trees [40]. It works by constructing a vast number of decision trees at training time and produce an individual tree model of classes (classification) or mean forecast (regression) as the end of product. Each tree is developed using a random subset of features employing an irregular subset of data that deviate the trees by appearing them diverse datasets. It has two parameters: the number of trees and the number of selected features at each node. DFR good in generate uneven data sets with missing variables since it is generally robust to overfitting. It also has lower classification errors and better f-scores than decision trees, but it does not easily to interpret the results [39]. Another disadvantage is the importance feature may not be vigorous to variety within the preparing dataset.

NNR is a chain of linear operations scattered with various nonlinear activation functions [41]. In general, the network has these defaults; the first layer is the input layer, the last layer is the output layer, and the hidden layer consisting of several nodes equal to the number of classes [42]. A neural network (NN) is defined by its structure, including the number of hidden layers, the number of nodes in each hidden layer, how the layers are linked, which activation function is used, and the weights. NN's are widely known for use in deep learning and modeling complex problems such as image recognition. They are easily adapted to regression problems. Thus, NNR is suited to situations where a more traditional regression model cannot fit a solution.

Bayesian Inference is used in the Bayesian approach, unlike linear regression [43]. Prior parameter information is combined with a likelihood function to generate parameter estimates, which means the forecast distribution evaluates the likelihood of a value y given x for a particular w, by means of likelihood by current belief about w given data (y, X). Finally, sum up all possible values of w [43]. BLR enables the survival of insufficient data or incorrectly distributed data by a fairly natural process. The major advantage is that, by this Bayesian processing, you recover the whole range of inferential solutions, rather than a point estimate and a confidence interval as in classical regression.

To summarize, the choice of the proposed methods to implement rainfall forecasting model is the difficulty for mimicking the rainfall process utilizing traditional model methods. This is due to the fact that the rainfall behavior is affected by different stochastic and natural resources such as temperatures rise and the air becomes warmer, more moisture evaporates from land and water to atmosphere and also climate change causes shifts in air and ocean currents which change the weather pattern.

2.3. Optimization technique

The optimization technique for M1 uses cross-validation with tuning. This optimization technique provides the best precision and helps to find problems with datasets by dividing data into some fold, then building and testing the models on each fold [44]. Numbers of input also played a role because the more input being generated, the highest result of coefficient of determination.

Meanwhile, the optimization technique for M2 uses varies types of normalization. Normalization is a method frequently applied as part of data preprocessing for ML. The purpose of normalization is to adjust the values of numeric columns in the dataset to use a standard scale, without distorting discrepancies in the spectrum of values or losing information [45]. It refers to scaling down the dataset so that the weight data lies between 0 and 1. This will help to easily compare corresponding normalized values from varies datasets that can eliminate the effects of large and smaller values of datasets. Basically, the normalization uses MinMax formula of $z = \frac{x - \min(x)}{[\max(x) - \min(x)]}$. In ML, this study has included several types of normalization which are MinMax, LogNormal and ZScore to optimize the modelling technique to get better result.

2.4. Scenarios

This study focuses on predicting the rainfall in different time horizons by using different ML algorithms. It has been divided into two methods with different scenarios to find the robust model mimicking the actual value. The methods are;

Table 3
Average and projected rainfall value with its' descriptive analysis.

Description	Mean (mm)	Standard Deviation (mm)	Minimum (mm)	Maximum (mm)	Median (mm)	Total of Daily Data
Average Value of 10 Stations (Actual)	10.04186	20.80274	0	271.3711	3.57	3,455
Projected Rainfall	8.523774	24.14694	0	788.22	1.38	3,455

2.4.1. Method 1: Forecasting rainfall using Autocorrelation Function (ACF)

Autocorrelation Function (ACF) is imperative analytical tools utilized with the time series analysis and forecasting [46]. Measuring the statistical relationships between observations is the primary purpose of these models in a single data series. ACF has a big advantage of measuring the amount of linear dependence between results of a time series which will be separated by a lag k. By plot these results along with the confidence band, it will define how well the present value of the series is related with its past values. ACF deals with all these components like trends, seasonality, cyclic, and residual while finding correlations; hence it called a complete auto-correlation plot. It is used to guess the form of the model and obtain approximate estimates of the parameters. Fig. 2(a to e) shows five scenarios of historical rainfall pattern in ACF.

For daily, the analysis took historical rainfall data from year 2010 until year 2019. Fig. 2(a) shows data observed for 31 days using January 2010 and shows that the first day of the month is more correlate compare to other days. Fig. 2(b) only observe for one year which is 2010 consisting of 52 weeks, so the graph only shows 52 of lag where 1 lag represents one week. Example lag 1 represent the beginning of 7 days for 2010 and continuously till lag 52 represents the last 7-days for 2010. Lag 1 to lag 4 show result for January 2010 while lag 48 to lag 52 show result for November 2010 and December 2010. Lag 1 to lag 4 and lag 52 give high results above the upper line where the amount of rainfall is high compared to other weeks.

Meanwhile, 10-days correlogram graphs (Fig. 2(c)) only observe for one year, year 2010, consisting 37 lags of 10 days. In the diagram, each lag equivalent to 10 days. For example, lag 1 represents the beginning of 10 days for year 2010 and continuously till lag 37 represent the last 10-days for year 2010. Lag 1 to lag 3 show result for January 2010 while lag 33 to lag 37 show result for November 2010 and December 2010. This lag gives high results above the upper line, where the amount of rainfall is high compared to other lag.

Fig. 2(d) show monthly rainfall data for year 2010 and year 2011. There are 24 of lag represent 24 months where lag 1 to lag 12 for 2011 and lag 13 to lag 24 for 2011. Lag 1 and 13 represent January, lag 11 and 23 represent November, lag 12 and 24 represent December. This correlogram shows that the data have seasonal dependencies and the same pattern over the years. For example, December 2011 (Lag 24) dependent on December 2010 (Lag 12), which means it is repeated. Based on the graph, there are dependencies between November, December and January in each year. Therefore, each year, the wet season occurs on November until January. But in some years, there are also dependencies between October and some years do not have dependencies in October. Therefore, in some years, the wet season may occur starting from October until January.

2.4.1.1. Input selection for method 1.

(i) Daily ACF

$$R_t = R_{t-1} \tag{2}$$

$$R_t + R_{t-1} = R_{t-2} \tag{3}$$

$$R_t + R_{t-1} + R_{t-2} = R_{t-3} \tag{4}$$

Refer to Fig. 2(a), there are three different scenarios is develop based on Lag 1, Lag 2 and Lag 3 shown in Eq. (2)–(4). This is because the autocorrelation function has more than 0.2. Therefore, Eq. (2) use R_t as input and R_{t-1} as output, where R_t is actual rainfall and R_{t-1} is rainfall at Lag 1. Eq. (3) use R_t and R_{t-1} as input and R_{t-2} as output, where R_{t-2} know and rainfall at Lag 2. Eq. (4) use R_t, R_{t-1} and R_{t-2} as input and R_{t-3} as output, where R_{t-3} know and rainfall at Lag 3.

(ii) Weekly ACF

$$R_t = R_{t-1} \tag{5}$$

$$R_t = R_{t-49} \tag{6}$$

$$R_t + R_{t-49} = R_{t-50} \tag{7}$$

Refer to Fig. 2(b), there are three different scenarios (Eqa. (5)–(7)) is develop based on Lag 1, Lag 49 and Lag 50 represent as week 1, week 49 and week 50 respectively. This is because the autocorrelation function has more than 0.2. Therefore, Eq. (5) use R_t as input and R_{t-1} as output, where R_t is actual rainfall and R_{t-1} is rainfall at Lag 1. Eq. (6) use R_t as input and R_{t-49} as output, where R_{t-49} know and rainfall at Lag 49. Eq. (7) use R_t and R_{t-49} as input and R_{t-50} as output, where R_{t-50} know and rainfall at Lag 50.

(iii) 10 Days ACF

$$R_t = R_{t-1} \tag{8}$$

$$R_t = R_{t-34} \tag{9}$$

$$R_t + R_{t-34} = R_{t-35} \tag{10}$$

$$R_t + R_{t-34} + R_{t-36} = R_{t-37} \tag{11}$$

Refer to Fig. 2(c), there are four different scenarios is develop based on Lag 1, Lag 34, Lag 35 and Lag 37 shown in Eq. (8)–(11). This is because the autocorrelation function has more than 0.2. Therefore, Eq. (8) use R_t as input and R_{t-1} as output, where R_t is actual rainfall and R_{t-1} is rainfall at Lag 1. Eq. (9) use R_t as input and R_{t-34} as output, where R_{t-34} know and rainfall at Lag 34. Eq. (10) use R_t and R_{t-34} as input and R_{t-35} as output, where R_{t-35} know and rainfall at Lag 35. Lastly Eq. (11) use R_t, R_{t-34}, R_{t-35} and R_{t-36} as input and R_{t-37} as output, where R_{t-37} know and rainfall at Lag 37.

(iv) Monthly ACF

$$R_t = R_{t-1} \tag{12}$$

$$R_t = R_{t-11} \tag{13}$$

$$R_t + R_{t-11} = R_{t-12} \tag{14}$$

$$R_t + R_{t-11} + R_{t-12} = R_{t-13} \tag{15}$$

Refer to Fig. 2(d), there are four different scenarios (Eq. (12)–(15)) is develop based on Lag 1, Lag 11, Lag 12 and Lag 13 represent as January 2010, November 2010, December 2010 and January 2011 respectively. This is because the autocorrelation function has more than 0.2. Therefore, Eq. (12) use R_t as input and R_{t-11} as output, where R_t is actual rainfall and R_{t-1} is rainfall at Lag 1. Eq. (13) use R_t as input and R_{t-11} as output, where R_{t-11} know and rainfall at Lag 11. Eq. (14) use R_t and R_{t-11} as input and R_{t-12} as output, where R_{t-12} know and rainfall at Lag 12. Lastly Eq. (15) use R_t, R_{t-11} and R_{t-12} as input and R_{t-13} as output, where R_{t-13} know and rainfall at Lag 13.

2.4.2. Method 2: Forecasting rainfall using projected error

This study aims to predict the error of projected rainfall to correct the future projected error. Data for this study include the projected rainfall consist of daily data from year 2010 until year 2099, which the projection point (in Fig. 3).

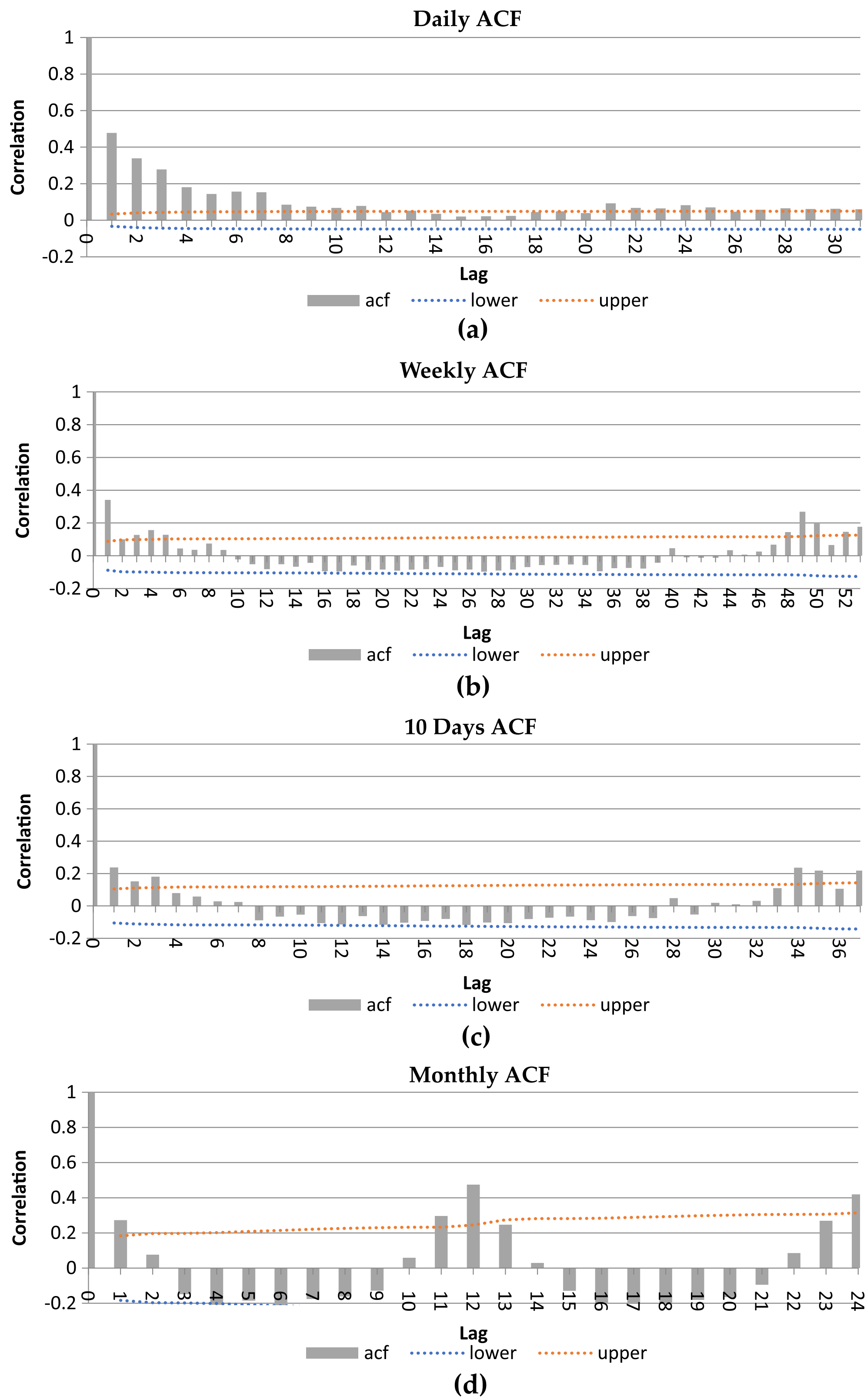

Fig. 2. AutoCorrelation Function (ACF) for; (a) Daily; (b) Weekly; (c) 10-days; and (d) Monthly.

The main input idea for forecasting rainfall using projected error is in Eq. (16). Where,R_p is projected rainfall as input for the model. As output for the model, E_p is the projected rainfall error. The main idea for this equation is to correct the projected rainfall from year 2020 to 2099 after predicting the projected rainfall error which the error is mainly from subtracted the projected rainfall R_p with actual rainfall R_a. The rainfall error data divided into several scenarios; daily, weekly, 10-days, and monthly.

$$R_p = E_p \tag{16}$$

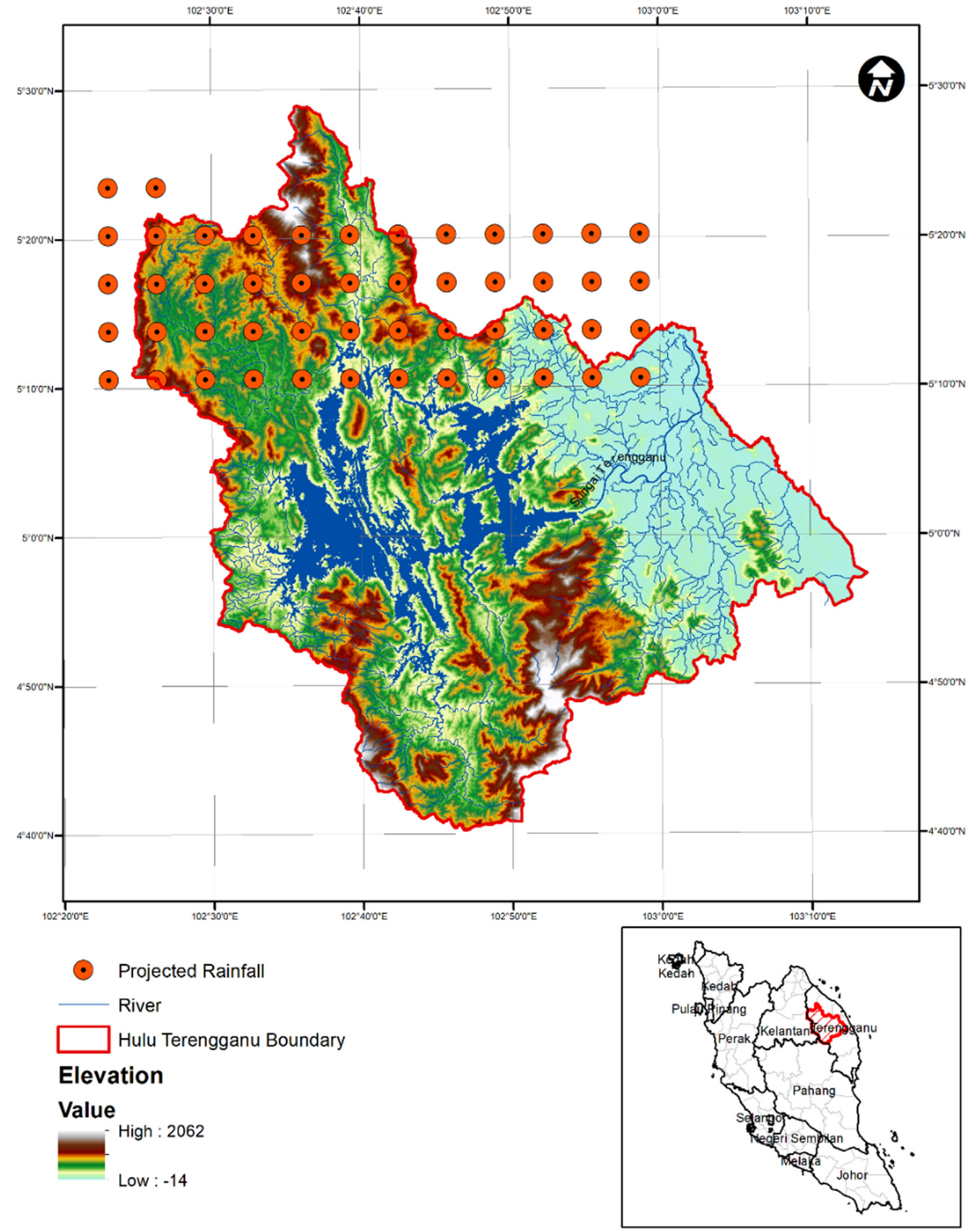

Fig. 3. Projected rainfall points.

2.5. Performance indicators

In this study, different model performance indicators were used as shown in Table 4 to signify the successful of scoring (datasets) has been by a trained model to mimicking the real values of the output parameters.

In a nutshell, the forecasting performance is better when the value of R^2 is close to 1 and differing for RMSE and MAE because the model's performance will be better if the value is close to 0 [50]. Fig. 4 showing the flow chart of the research methodology.

3. Results and discussion

3.1. Method 1: Forecating rainfall using Autocorrelation Function (ACF)

Table 5 shows the best models results to predict rainfall based on ACF. In this study, there are four different regression has been used, which is Bayesian Linear Regression (BLR), Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR) and Neural Network Regression (NNR). All regression generates different scenario for rainfall data divided into daily, weekly, 10-days and

Table 4
Performance indicators to evaluate the model.

Equation	Performance indicator
$MAE = \frac{1}{n}\sum_{i=1}^{N}\lvert MSL_p - MSL_o\rvert$	Mean absolute error, MAE [47] reflects the degree of absolute error between the actual and forecasted data.
$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(MSL_p - MSL_o)^2}{N}}$	Root Mean Square Error, RMSE [47] is compared between the actual and forecasted data.
$RAE = \frac{\sum_{i=1}^{n}\lvert p_i - a_i\rvert}{\sum_{i=1}^{n}\lvert \bar{a} - a_i\rvert}$	Relative absolute error, RAE [48] is the relative absolute difference between actual and forecasted values.
$RSE = \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{\sum_{i=1}^{n}(\bar{a} - a_i)^2}$	Relative squared error, RSE [48] similarly normalizes the entire squared error of the forecasted values.
$R^2 = \frac{\sum_{i=1}^{n}\left(MSL_o - \bar{MSL_o}\right)(MSL_p - \bar{MSL_p})}{\sqrt{\sum_{i=1}^{n}(MSL_o - \bar{MSL_o})^2\sum_{i=1}^{n}(MSL_p - \bar{MSL_p})^2}}$	Coefficient of determination, R^2 [49] is showing the performance of forecasting model where zero means the model is random while 1 means there is a perfect fit.

monthly. Hence, the best regression developed for ACF is BDTR since it has the highest coefficient of determination, R2.

It can be seen from the table that without tuning, the BDTR model did not perform well, whereas, noticeable improvement in the accuracy of the proposed model can be seen after adding cross validation and hyper-parameters tuning techniques. By adding this step into the ML, the proposed BDTR model gives the best accuracy in predicting the rainfall where the value of coefficient of determination in predicting daily rainfall it ranges between (0.5525075-0.9739693), and for weekly rainfall prediction it ranges between (0.8400668-0.989461), and for 10 days rainfall prediction it ranges between (0.8038288-0.9894429) and finally for monthly rainfall prediction it ranges between (0.9174191-0.9998085). Based on these results, it can be concluded that BDTR capable of predicting the rainfall in different time horizon with acceptable level of accuracy and the accuracy of the model improved when more inputs included.

3.2. Method 2: Forecasting rainfall using projected error

Table 6 summarizes the metrics for each top three best models for different scenarios using different normalization and data partitioning. In this context, different normalization such as ZScore, LogNormal, and MinMax and data partitioning (80% and 90%) are investigated to obtain the optimal model with high level of accuracy. Overall model performances show that normalization using LogNormal gives a good result for each category except for 10-days prediction. Comparison between four models used, BDTR

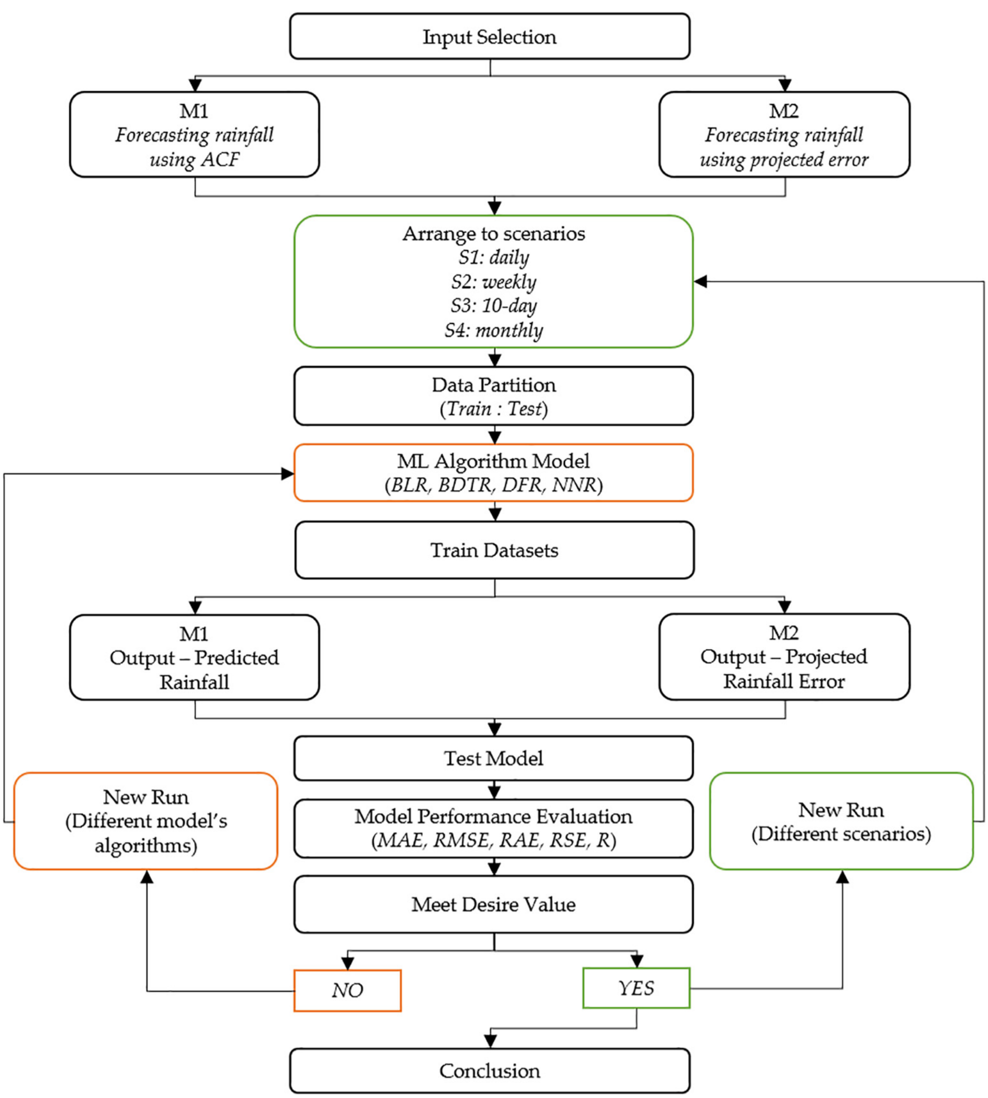

Fig. 4. The methodology of current research.

Table 5
Result for the best model in M1 using ACF.

Scenario	Regression	Model	Coefficient of Determination
(a) Daily			
$R_t = R_{t-1}$	BDTR	Without tuning	0.2458173
		With tuning	0.5525075
$R_t + R_{t-1} = R_{t-2}$	BDTR	Without tuning	−0.1383447
		With tuning	0.8468193
$R_t + R_{t-1} + R_{t-2} = R_{t-3}$	BDTR	Without tuning	−0.2020856
		With tuning	0.9739693
(b) Weekly			
$R_t = R_{t-1}$	BDTR	Without tuning	−0.0002462
		With tuning	0.8400668
$R_t = R_{t-49}$	BDTR	Without tuning	−0.1179256
		With tuning	0.8825647
$R_t + R_{t-49} = R_{t-50}$	BDTR	Without tuning	0.235989
		With tuning	0.989461
(c) 10 Days			
$R_t = R_{t-1}$	BDTR	Without tuning	−1.0041807
		With tuning	0.8038288
$R_t = R_{t-34}$	BDTR	Without tuning	0.1182632
		With tuning	0.8949389
$R_t + R_{t-34} = R_{t-35}$	BDTR	Without tuning	−0.4616042
		With tuning	0.9607741
$R_t + R_{t-34} + R_{t-35} + R_{t-36} = R_{t-37}$	BDTR	Without tuning	−0.1377565
		With tuning	0.9894429
(d) Monthly			
$R_t = R_{t-1}$	BDTR	Without tuning	0.1163886
		With tuning	0.9174191
$R_t = R_{t-11}$	BDTR	Without tuning	0.0514856
		With tuning	0.6941756
$R_t + R_{t-11} = R_{t-12}$	BDTR	Without tuning	−0.5693955
		With tuning	0.9939951
$R_t + R_{t-11} + R_{t-12} = R_{t-13}$	BDTR	Without tuning	−0.4366365
		With tuning	0.9998085

and DFR, are the most acceptable result than NNR and BLR. The outcome of the results shows the best model for daily error prediction with R equal to 0.737978 is BDTR and for weekly rainfall error prediction where R equal to 0.7921. While for monthly rainfall error prediction, DFR outperformed other models where R equal to 0.7623. However, for 10-days rainfall error prediction, NNR model with ZScore normalization outperformed other models in predicting the value of 10-days with acceptable level of accuracy where R is equal to 0.61728. It can be concluded, acceptable level of accuracy could be achieved by reducing the error in the dataset of the projected rainfall with the expected observable rainfall by using BDTR integrated with LogNormal and by partitioning the data to 90% for training and 10% for testing.

Finally, Fig. 5 demonstrates the actual error and the predicted error. The figure shows how well the proposed model can resemble the actual error between the observed and projected rainfall during the testing phases. The red line demonstrates the predicted value from the proposed ML algorithm, while blue line demonstrates the observed value of actual error. It can be seen that the proposed model has acceptable level of accuracy for all four different time horizons, however, highest level of accuracy obtained for the weekly error (Fig. 5(b).

4. Conclusion

This paper focuses on two methods; (1) Forecasting rainfall using Autocorrelation Function (ACF) based on the historical rainfall data and (2) Forecasting rainfall using Projected Error based on historical and projected rainfall data. Both methods using different algorithms such as BDTR, DFR, BLR and NNR to identify the optimal prediction for rainfall and different time horizons. The results presented that for M1, the result gets better with cross-validation with BDTR and tuning its parameter. The more input included to the model; the more accurate the model can perform. The best regression developed for ACF is BDTR since it has the highest coefficient of determination, R2 (daily: 0.5525075, 0.8468193, 0.9739693; weekly: 0.8400668, 0.8825647, 0.989461; 10 days: 0.8038288, 0.8949389, 0.9607741, 0.9894429; and monthly: 0.9174191, 0.6941756, 0.9939951, 0.9998085) meaning the better rainfall prediction for the future. For method 2, a variation of result when using different normalization techniques and shows using LogNormal normalization with BDTR and DFR gives the best model mimicking the actual projected error. The best-predicting scenarios is

Table 6
Summary of top best three models for each scenario.

Regression	Normalization	Training (%)	MAE	RMSE	RAE	RSE	R
(a) Daily Error							
BDTR	LogNormal	90	0.105467	0.160547	0.391675	0.262022	0.737978
DFR	LogNormal	90	0.106306	0.177975	0.394791	0.321997	0.678003
BDTR	LogNormal	80	0.113675	0.178906	0.432401	0.332167	0.667833
(b) Weekly Error							
BDTR	LogNormal	90	0.064627	0.117037	0.314988	0.2079	0.7921
DFR	LogNormal	90	0.058325	0.126458	0.284272	0.242716	0.757284
DFR	LogNormal	80	0.078724	0.160686	0.348225	0.349252	0.650748
(c) 10-Days Error							
NNR	ZScore	80	0.389449	0.672152	0.716514	0.38272	0.61728
BLR	ZScore	80	0.417034	0.674603	0.767265	0.385516	0.614484
DFR	MinMax	80	0.031469	0.053312	0.801608	0.461544	0.538456
(d) Monthly Error							
DFR	LogNormal	80	0.094573	0.156451	0.379418	0.2377	0.7623
DFR	LogNormal	90	0.151092	0.219738	0.47616	0.367674	0.632326
DFR	MinMax	90	0.048869	0.069285	0.609473	0.372461	0.627539

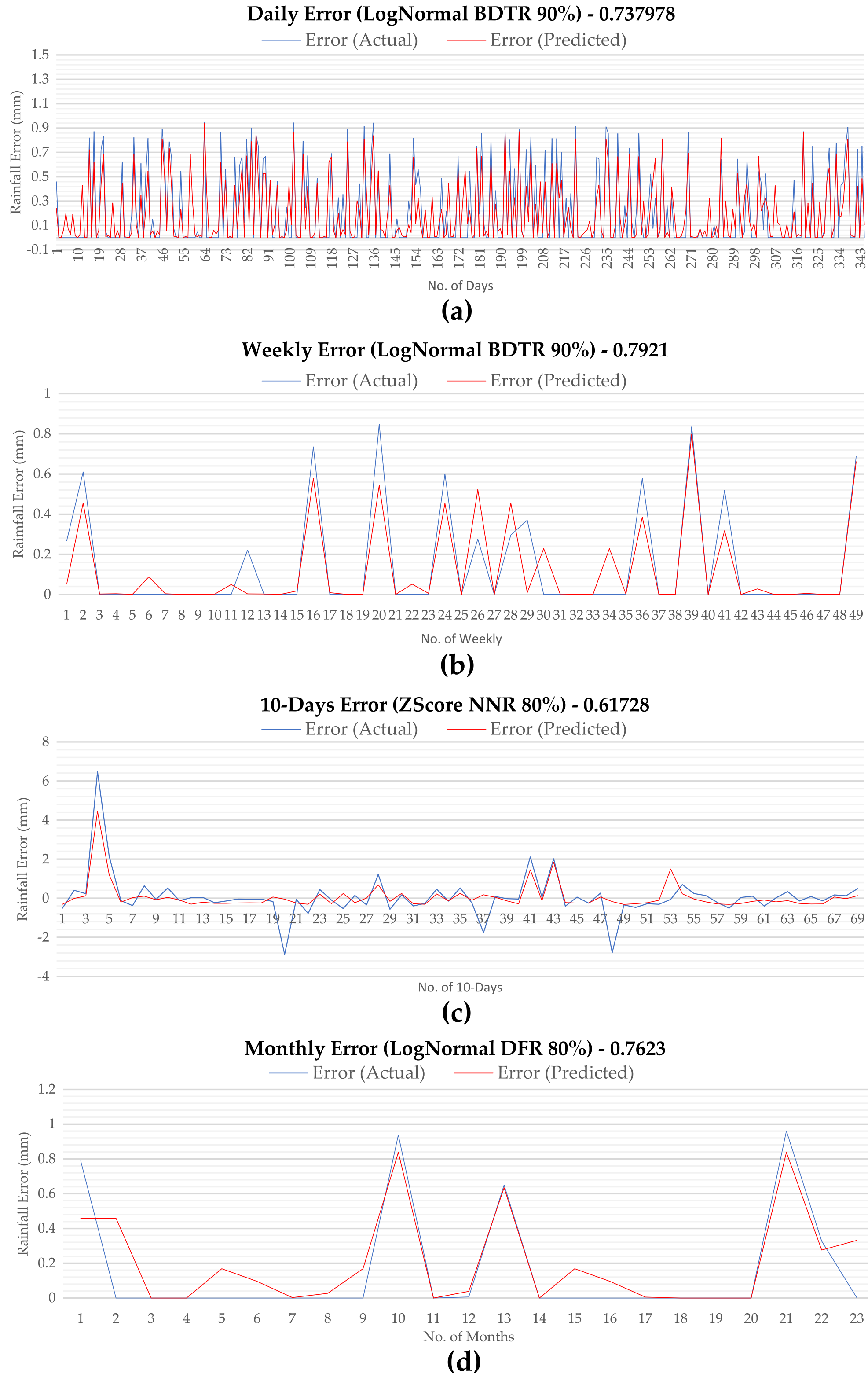

Fig. 5. Comparison between actual error and predicted error for; (a) Daily; (b) Weekly; (c) 10-days; and (d) Monthly.

the weekly error with R of 0.7921 which is the closest to 1. This means the BDTR model can best predict weekly average error to correct the weekly average projected rainfall. In conclusion, method 1 is the best prediction for the rainfall, mimicking the actual values with the highest coefficient closer to 1. The dependencies on ACF show that rainfall has almost a similar pattern every year from November to January, and this shows a correlation between the predicting input and output. The findings of the current study showed that standalone machine learning algorithms capable to predict the rainfall with acceptable level of accuracy, however, more accurate rainfall prediction might be achieved by proposing hybrid machine learning algorithms and with the inclusion of different climate change scenarios.

Funding

This research and the APC was funded by TNB Seed Fund (U-TG-RD-19-01) managed by UNITEN R&D Sdn Bhd.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Asset Management Department, Generation Division, Tenaga Nasional Berhad, Malaysia for providing this study with the rainfall data and permission to conduct the study and the authors would like to acknowledge the financial support received from TNB Seed Fund (U-TG-RD-19-01) managed by UNITEN R & D Sdn. Bhd. In addition to that, the authors would like to thank Professor Dr.Ir.Lariyah Mohd Sidek for providing the projected rainfall data which used in this study.

References

[1] Ismail M, Suroto A, Abdullah S, Nagarani N, Arumugam K, Kumaraguru VJ, et al. Response of Malaysian local rice cultivars induced by elevated ozone stress EnvironmentAsia genotoxicity assessment of mercuric chloride in the Marine Fish Therapon Jaruba.

[2] Abdullah S, Ismail M. The weather and climate of tropical Tasik Kenyir, Terengganu. In: Greater Kenyir landscapes: social development and environmental sustainability: from ridge to reef. Springer International Publishing: Cham; 2019. p. 3–8. doi:10.1007/978-3-319-92264-5_1.

[3] Khairul Amri Kamarudin M, Abd Wahab N, Juahir H, Mohd Firdaus Nik Wan N, Barzani Gasim M, et al. The potential impacts of anthropogenic and climate changes factors on surface water ecosystem deterioration at Kenyir Lake, Malaysia. Int J Eng Technol 2018; 7 (3.14): 67. doi:10.14419/ijet.v7i3.14.16864.

[4] Chang TK, Talei A, Alaghmand S, Ooi MPL. Choice of rainfall inputs for event-based rainfall-runoff modeling in a catchment with multiple rainfall stations using data-driven techniques. J Hydrol 2017;545:100–8. doi: https://doi.org/10.1016/j.jhydrol.2016.12.024.

[5] Chang TK, Talei A, Quek C, Pauwels VRN. Rainfall-runoff modelling using a self-reliant fuzzy inference network with flexible structure. J Hydrol 2018;564:1179–93. doi: https://doi.org/10.1016/j.jhydrol.2018.07.074.

[6] Bartoletti N, Casagli F, Marsili-Libelli S, Nardi A, Palandri L. Data-driven rainfall/runoff modelling based on a neuro-fuzzy inference system. Environ Model Softw 2018;106:35–47. doi: https://doi.org/10.1016/j.envsoft.2017.11.026.

[7] Ghumman AR, Ghazaw YM, Sohail AR, Watanabe K. Runoff forecasting by artificial neural network and conventional model. Alexandria Eng J 2011;50 (4):345–50. doi: https://doi.org/10.1016/j.aej.2012.01.005.

[8] Climate changes impacts towards sedimentation rate at Terengganu River, Terengganu, Malaysia | J Fundam Appl Sci https://www.ajol.info/index.php/jfas/article/view/168262 (accessed Jun 12, 2020).

[9] (PDF) Multiple Linear Regression (MLR) models for long term Pm10 concentration forecasting during different monsoon seasons https://www.researchgate.net/publication/318777794_Multiple_Linear_Regression_MLR_models_for_long_term_Pm10_concentration_forecasting_during_different_monsoon_seasons (accessed Mar 27, 2020).

[10] Peredaran S; Dan A, Monsun S, Daya B, Perairan D, Timur P, et al. Simulation of Southwest Monsoon Current Circulation and Temperature in the East Coast of Peninsular Malaysia 2014; 43.

[11] Duan A, Publisher Q. A global optimization strategy for efficient and effective calibration of hydrologic models. Item Type Text; Dissertation-Reproduction (Electronic). The University of Arizona. 1991.

[12] Luk KC, Ball JE, Sharma A. An application of artificial neural networks for rainfall forecasting. Math Comput Model 2001;33(6–7):683–93. doi: https://doi.org/10.1016/S0895-7177(00)00272-7.

[13] Duan Q, Sorooshian S, Gupta VK. Optimal use of the SCE-UA global optimization method for calibrating watershed models. J Hydrol 1994;158 (3–4):265–84. doi: https://doi.org/10.1016/0022-1694(94)90057-4.

[14] Box G, Jenkins G, Reinsel G, Ljung G. Fifth edition time series analysis forecasting and control. In: Balding D, Cressie N, Fitzmaurice G, Givens G, Goldstein H, Molenberghs G, Scott D, Smith A, Tsay R, Weisberg S, editors. John Wiley & Sons; 2016.

[15] Tokar AS, Markus M. Precipitation-runoff modelling using artificial neural networks and conceptual models. J Hydrol Eng 2000;5(April):156–61.

[16] Najah AA, El-Shafie A, Karim OA, Jaafar O. Integrated versus isolated scenario for prediction dissolved oxygen at progression of water quality monitoring stations. Hydrol Earth Syst Sci Discuss 2011;8(3):6069–112.

[17] Hipni A, El-shafie A, Najah A, Karim OA, Hussain A, Mukhlisin M. Daily forecasting of dam water levels: comparing a Support Vector Machine (SVM) model with Adaptive Neuro Fuzzy Inference System (ANFIS). Water Resour Manag 2013;27(10):3803–23. doi: https://doi.org/10.1007/s11269-013-0382-4.

[18] Mohammadi B, Ahmadi F, Mehdizadeh S, Guan Y, Pham QB, Linh NTT, et al. Developing novel robust models to improve the accuracy of daily streamflow modeling. Water Resour Manag 2020.

[19] Mohammadi B, Linh NTT, Pham QB, Ahmed AN, Vojteková J, Guan Y, et al. Adaptive neuro-fuzzy inference system coupled with shuffled frog leaping algorithm for predicting river streamflow time series. Hydrol Sci J 2020:1.

[20] Mohammadi B, Mehdizadeh S. Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. Agric Water Manag 2019;2020(237):106145.

[21] Mohammadi B, Aghashariatmadari Z. Estimation of solar radiation using neighboring stations through hybrid support vector regression boosted by krill herd algorithm. Arab J Geosci 2020;13(10):1–16.

[22] Moazenzadeh R, Mohammadi B. Assessment of bio-inspired metaheuristic optimisation algorithms for estimating soil temperature. Geoderma 2018;2019(353):152–71.

[23] Muslim TO, Ahmed AN, Malek MA, Afan HA, Ibrahim RK, El-shafie A, et al. Investigating the influence of meteorological parameters on the accuracy of sea-level prediction models in Sabah. Malaysia 2020.

[24] French MN, Krajewski WF, Cuykendall RR. Rainfall forecasting in space and time using a neural network. J Hydrol 1992;137(1–4):1–31. doi: https://doi.org/10.1016/0022-1694(92)90046-X.

[25] Luk KC, Ball JE, Sharma A. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. J Hydrol 2000;227(1–4):56–65. doi: https://doi.org/10.1016/S0022-1694(99)00165-1.

[26] Valverde Ramírez MC, De Campos Velho HF, Ferreira NJ. Artificial neural network technique for rainfall forecasting applied to the São Paulo Region. J Hydrol 2005;301(1–4):146–62. doi: https://doi.org/10.1016/j.jhydrol.2004.06.028.

[27] Bari SH, Rahman MT, Hussain MM, Ray S. Forecasting monthly precipitation in Sylhet city using ARIMA model. Civil Environ. 2015;7:1.

[28] Yu PS, Yang TC, Chen SY, Kuo CM, Tseng HW. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. J Hydrol 2017;552:92–104. doi: https://doi.org/10.1016/j.jhydrol.2017.06.020.

[29] Kumar S, Roshni T, Himayoun D. A comparison of Emotional Neural Network (ENN) and Artificial Neural Network (ANN) approach for rainfall-runoff modelling. Civ Eng J 2019;5(10):2120–30. doi: https://doi.org/10.28991/cej-2019-03091398.

[30] Jafari Nodoushan E. Monthly forecasting of water quality parameters within bayesian networks: a case study of Honolulu, Pacific Ocean. Civ Eng J 2018;4 (1):188. doi: https://doi.org/10.28991/cej-030978.

[31] Sumi SM, Zaman MF, Hirose H. A rainfall forecasting method using machine learning models and its application to the fukuoka city case. Int J Appl Math Comput Sci 2012;22(4):841–54. doi: https://doi.org/10.2478/v10006-012-0062-1.

[32] Nakagawa S, Freckleton RP. Missing inaction: the dangers of ignoring missing data. Trends Ecol Evol 2008;23(11):592–6.

[33] Jollife IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans R Soc A Math Phys Eng Sci 2016;374(2065).

[34] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemom Intell Lab Syst 1987;2(1–3):37–52.

[35] Lal BB, Al-mashidani G. A Technique for the determination of areal average rainfall/Une Méthode Pour La Détermination de Précipitation Moyenne Pour Une Zone Donnée a technique for the determination of areal average rainfall. Hydrol Sci Sci Hydrol 1978;23. doi: https://doi.org/10.1080/02626667809491823.

[36] Boosted Decision Tree Regression – Azure Machine Learning Studio | Microsoft Docs https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/boosted-decision-tree-regression (accessed Oct 23, 2019).

[37] Boosted Trees Regression | Turi Machine Learning Platform User Guide https://turi.com/learn/userguide/supervised-learning/boosted_trees_regression.html (accessed Oct 23, 2019).

[38] Pros and cons of classical supervised ML algorithms - rmartinshort https://rmartinshort.jimdofree.com/2019/02/24/pros-and-cons-of-classical-supervised-ml-algorithms/ (accessed Feb 26, 2020).

[39] HackingNote https://www.hackingnote.com/en/machine-learning/algorithms-pros-and-cons (accessed Feb 26, 2020).

[40] Criminisi A. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found Trends® Comput Graph Vis 2011;7(2–3):81–227. doi: https://doi.org/10.1561/0600000035.

[41] Tan LK, McLaughlin RA, Lim E, Abdul Aziz YF, Liew YM. Fully automated segmentation of the left ventricle in cine cardiac MRI using neural network regression. J Magn Reson Imaging 2017;48(1):140–52.

[42] Damian DC. A critical review on artificial intelligence models in hydrological forecasting how reliable are artificial intelligence models. 2019, No. July.
[43] Catal C, Ece K, Arslan B, Akbulut A. Benchmarking of regression algorithms and time series analysis techniques for sales forecasting. Balk J Electr Comput Eng 2019;7.
[44] Tune Model Hyperparameters – ML Studio (classic) – Azure | Microsoft Docs https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/tune-model-hyperparameters (accessed Mar 31, 2020).
[45] Normalize Data: Module Reference – Azure Machine Learning | Microsoft Docs.
[46] Parmar A, Mistree K, Sompura M. Machine learning techniques for rainfall prediction: a review indian sign language recognition view project machine learning view project machine learning techniques for rainfall prediction: a review.
[47] Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast 2006;22(4):679–88.
[48] Miller TB, Kane M. The precision of change scores under absolute and relative interpretations. Appl Meas Educ 2001;14(4):307–27.
[49] Nagelkerke NJD. A note on a general definition of the coefficient of determination. Biometrika 1991;78(3):691–2.
[50] Cheng CT, Feng ZK, Niu WJ, Liao SL. Heuristic methods for reservoir monthly inflow forecasting: a case study of Xinfengjiang reservoir in pearl river, China. Water (Switzerland) 2015;7(8):4477–95.