Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Analysing Hydrological Extremes with Neural Network Algorithm and Support Vector Machine

Nur Afiqah Abdul Rahim[1], Puteri Natasha Sofia Zulkafli [2], Shuhaida Ismail[1,*], Azme Khamis[2]

[1] Faculty of Computer Science & Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400, Malaysia
[2] Faculty of Applied Science and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh, 86400, Malaysia

ABSTRACT

Many areas in Malaysia that have never experienced flooding have been affected by the floods over the past few years. As a result of these events, water levels can change abruptly, resulting in extreme hydrological data values. This study investigated the characteristics of daily rainfall amounts at the Petaling Jaya station using two statistical tests, the Mann-Kendall test and Anderson-Darling test. After that, two machine learning models, namely Artificial Neural Networks and Support Vector Machines, were tested to predict and analyze the hydrological data. These investigations assessed the impact of multiple hidden nodes on rainfall amounts in ANN. In addition, they assessed performance accuracy based on the use of different kernels in SVM. The results of four performance metrics were examined, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of correlation (R) to demonstrate that ANNs are more accurate than SVMs with respect to prediction accuracy. In addition to stable prediction, ANN can capture extreme events more effectively than SVM.

Keywords:
Artificial neural network; support vector machine; extreme value; hydrology; rainfall forecasting

1. Introduction

Hydrology is the study of water movement, distribution, and quality on Earth. The field is critical because it aids an understanding of the water cycle, water resources availability, and climate change effects. Managing resources, understanding the water cycle, and predicting climate change impacts require hydrological data. Climate change affects many aspects of the natural water management system. These aspects include the elements and mechanisms of precipitation-runoff, as well as the amplitude, intensity, and river flow during floods and droughts, which affect peak flow hydrograph events during particular times. Future impacts of these factors are likely to persist, and they may negatively impact freshwater availability and sustainability as well as the natural environment [1]. It is crucial to understand precipitation trends when developing and managing water reserves. This is especially important given new evidence of hydrological extreme events and their unpredictability. It is one of the primary reasons for natural accidents, for instance, when massive flooding occurs in a

* *Corresponding author.*
E-mail address: shuhaida@uthm.edu.my

https://doi.org/10.37934/araset.61.2.112

particular area. This causes financial harm, extensive property destruction, and sometimes fatalities [2,3].

Extreme events such as storms have a profound impact on how hydrologic systems are characterized and water infrastructure is designed in hydrology. Hydrology remains challenged to capture and describe extremes in natural systems due to their complexity and variability. Extreme Value Theory (EVT) is commonly applied to studying extreme values in hydrological systems [4, 5]. Noise or missing values can greatly impact a model's precision due to the quality of the input data. Pre-processing input data with Singular Spectrum Analysis (SSA) and hybrid Artificial Neural Network and Empirical Mode Decomposition (EMD) models can enhance precision [6]. The choice of a model considers elements during decision making, such as data type, problem difficulty, computing resources, and insights needed. Ensemble techniques, hybrid models, and domain knowledge integration can improve prediction capability and interpretability in hydrological abnormalities analyses [7].

It has become increasingly popular over the past decade to use machine learning approaches to model hydrological processes and predict hydrological variables based on historical hydrological data [8-11]. There has been a significant increase in the popularity of machine learning techniques over the last few years, including Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and many others based on historical data [3, 12, 13]. In the advanced technology era, the use of ANN is a common method of implementing macro data into daily activities. Due to their ability to learn and generalize from enough data pairs, artificial neural networks are capable of handling large-scale and difficult problems.

ANNs are based on biological neurons in the human brain [14]. Although it is impossible to directly imitate biological neuron functions, it is still able to solve a wide variety of complicated problems. In ANNs, neurons are arranged in layers, with input, hidden, and output layers, with neurons from one layer connecting to neurons from the next layer [15-17]. Data flow begins with the input layer, which accepts data and forwards it to the next node. The data then performs all the complicated procedures in the hidden layer. A hidden layer may be present in each issue based on the needs of the issue. The results are stored in the output layer, which is the last layer. The neural network is generally capable of generating the best outcome without modifying the input data in any way. In addition to offering adequate precision in hydrological research, this type of system also provides more durable equipment for engineering applications than traditional approaches.

ANNs have previously been shown to be the most popular machine learning model in hydrology. ANN has demonstrated its capability in predicting a broad range of hydrological variables, including stream flow, groundwater level, and water quality [17]. The results of a previous study [18], showed that ANNs could provide reliable predictions of stream flow, especially when combined with other models, such as physical models or other machine learning models. As a result, the ANN proved to be superior in modelling hourly rainfall-runoff data as well as being able to establish the relationship between rainfall and runoff [19].

Additionally, numerous studies have demonstrated that nonlinear hydrologic processes can be effectively modelled using ANNs, such as precipitation, runoff from rain, river flows, groundwater resources, water integrity models, and streams. According to Banhatti *et al.,* [20], the ANN model is sufficiently robust to change dramatically when the training data is changed, which indicates it is sensitive to the analysed data. Moreover, the effectiveness of the model has a significant impact on training procedures, model design, and data pre-processing approaches. Through a downscaling model, ANNs could be used to predict flood peak discharges and seasonal stream flows. Other than that, utilizing ANN ensembles, stream flow conditions were predicted using a downscaling model [21]. They also found that ANNs for hydrological prediction had increased in recent years, with many studies

showing promising results.

An ANN prediction's accuracy depends on many factors, including the choice of input variables, the design of the model, and the training and testing techniques. As a result of the nature of ANNs, they are often called "black-box" models, which pose many challenges when it comes to creating predictions. Aside from this, ANNs require a greater amount of training time than other models. To overcome the limitations of ANNs, a new method was developed called Support Vector Machines (SVM).

SVM has been demonstrated to be superior in many research areas, including image classification, financial forecasting, medical diagnosis, and many others. SVMs are powerful tools that are efficient and effective at solving complex problems. In addition to its ability to identify patterns in data quickly, it can also assist in making decisions based on that data. SVM is also able to learn from new data and adjust its predictions accordingly. There has been a lot of interest in SVM in hydrological studies. SVM has been frequently used in modelling runoff forecasting due to their advantages, which include theoretical global optimization, avoiding dimensionality catastrophes, and large sample sizes [22]. Previous study by Han *et al.,* [23] proposed SVM implementation in flood forecasting. The study revealed a fascinating result in SVM with respect to different rainfall volumes. According to Hirani *et al.,* [24], SVM is more reliable than other forecasting approaches, both statistical and numerical, in predicting rainfall values. Although SVM models are effective in achieving better classification and prediction results, they are also subject to some limitations. These limitations include the fact that their accuracy is strongly influenced by parameter settings and kernel functions [25].

Since hydrological data is often nonlinear and nonstationary, this research aims to prove that hydrological characteristics can be demonstrated using Mann-Kendall and Anderson Darling tests. The next step is to test the capability of ANNs and SVMs to analyse extreme values within hydrological datasets. In this study, ANNs and SVMs are employed with a variety of hidden layers and kernel types. Finally, comparative analyses were done based on several performance measurements. This study provides insight into the analysis of hydrological extreme values and the best model to deal with them.

2. Methodology

In the methodology part, the methods and material used in the study will be discussed and explained in detail.

2.1 Data Collection

The data was acquired from the Malaysian Meteorological Department (Met Malaysia) and represents hydrological data for the Petaling Jaya, Selangor region. In this study, the daily rainfall amount (mm) variable from 2010 to 2020 is used to forecast the daily rainfall amount (mm) for 2021. There are eight variables including year, month, day, daily maximum, and minimum temperatures in degrees Celsius, daily relative humidity in percentages, daily rainfall totals in millimeters and daily mean sea level pressure in hPa. Total observations of 5844 data points in this dataset. The accuracy of the forecast was evaluated between the optimal selection of hidden layer to determine which model could provide the greatest accuracy. Figure 1 showed the time series plot of daily rainfall used in this study. The time series plot shows a high fluctuation up-down indicating seasonal and irregular variations.

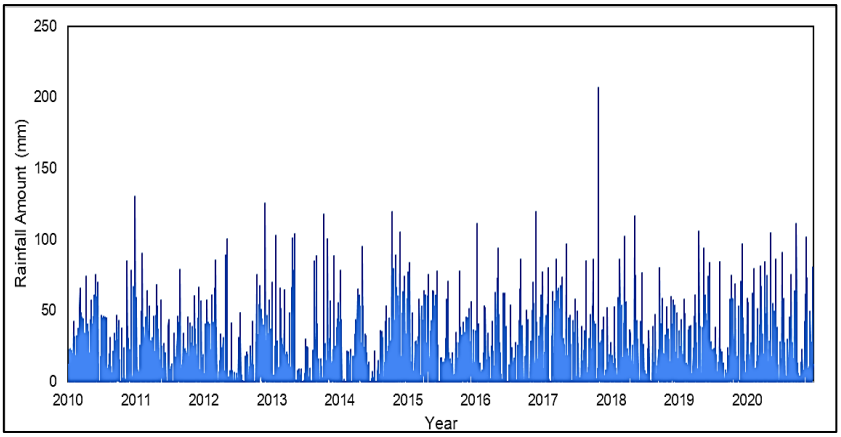

Fig. 1. Time series of daily rainfall amount in mm for 11 years

2.2 Preliminary Test

Preliminary tests are run to make sure the data meet the assumptions before moving further with the experiment. The purpose of the preliminary tests is to examine the dataset's trend existence, and normality. The Mann-Kendall, and Anderson Darling tests have been used in this investigation.

2.2.1 Mann-Kendall test

The Mann-Kendall test is a nonparametric method applied for recognition monotonic patterns in time series data. Monotonic trends refer to trends where the direction of the trend is consistent, but the magnitude of the trend may not be constant. The test statistic for the Mann-Kendall test is the Kendall's tau statistic, which determines the strength of the relationship between the data and a monotonic function. To perform the test, the data are first ranked, and then the difference in ranks between each pair of observations is computed. The test statistic is then computed as the sum of the signed differences. A positive tau statistic indicates an upward trend, a negative tau statistic indicates a downward trend, and a tau statistic of zero indicates no trend. The null hypothesis for the test is that there is no monotonic trend in the data.

2.2.2 Anderson Darling test

The Anderson-Darling test is a statistical test that can be used to check if a sample of data comes from a specific probability distribution, most commonly the normal distribution. The test statistic for the Anderson-Darling test is calculated by comparing the sample cumulative distribution function (CDF) with the theoretical CDF of the normal distribution. The null hypothesis for the Anderson-Darling test is that the sample comes from a normal distribution. If the calculated test statistic is larger than a critical value from the Anderson-Darling table, the null hypothesis is rejected, and it suggests that the data does not come from a normal distribution.

2.3 Pre-Processing Data

Data screening and cleansing were done at the pre-processing stage. Boxplots were plotted to show outliers, while histograms with normal curves are used to verify normal distributions. In addition, the quantity of missing data, kurtosis, and skewness are computed and analyzed in the statistics descriptive table. Although, the outliers are left in the data since it was used to identify

extreme values. The data was split into 20% of the dataset for testing set while another 80% of the dataset for testing set. Therefore, training datasets were normalized to transform the data into smaller and appropriate range using Minimum-Maximum scaler.

2.4 Model Development using ANN

An Artificial Neural Network (ANN) mimics the basic activities of neurons in the human brain by simulating the core functions of a biological neuron. Based on robust computational power, artificial intelligence algorithms mimic the operation of biological neurons that can solve a wide range of problems. It is apparent that the exceptional computational resilience of ANN algorithms is the source of their intelligence.

In these neural networks, three layers are involved, including input layers, hidden layers, and output layers. Input layer nodes define the independent variables, whilst output layer nodes define the dependent variable. Processing is performed by each node and each node's synaptic weight describes the amount it contributes to the overall output. As a nonlinear functional mapping of the explanatory variable is performed between each node, the relationship between them can be described as,

$$X(t) = b + \sum_{i=1}^{nn} w_i \times f(\sum_{k=1}^{L} W_{ik} \times X(t-k) + B \quad (1)$$

where the output is the forecast value, ***X(t)*** , input variables, $X(t-k)$ and k = 1, 2, . . ., *L* , synaptic weight which relates hidden layer nodes , w_i , output layer node, i , synaptic weight which relates input layer nodes, W_{ik}, activation function , f , bias factors that correspond to the output is B while bias factors that correspond to the hidden layer is b. Table 1 showed the hyperparameter setting for ANN model used in this study. Based on Table 1, the activation function used in this study was "*logistic*", which refers to the logistic sigmoid function, and the solver used was "*lbfgs*", otherwise known as an optimization method in the family of quasi-Newton methods. The alpha value, batch size, primary learning rate, power *t*, and momentum value applied in this study are 0.0001, 32, 0.001, 0.5, and 0.9 respectively which is the default of the values. The learning rate used in the study is "*adaptive*" in which the learning rate remains constant as long as the training loss declines. The maximum number of iterations, also known as the maximum number of epochs used in the study is 100.

Table 1
Hyperparameter of ANN model

Parameter	Value
	ANN
Activation	Logistic
Solver	lbfgs
Alpha	0.0001
Batch size	32
Learning rate	Adaptive
Initial learning rate	0.001
Power t	0.5
Maximum epochs	100
Momentum	0.9

Table 2 displays the value for each hidden layer selected. For the selection of hidden layers, it is advisable that the data with more complex structure apply a bigger size of hidden layer to obtain an

optimal solution. In this research, only one hidden layer is used with several selection of hidden layer nodes which are n/2, *n*+3, 2*n* and 2*n*+1.

Table 2
Number of hidden layers

Hidden layer	Nodes
n/2	4
n+3	10
2n	14
2n+1	15

2.5 Model Development using Support Vector Machine (SVM)

SVM is a previously developed machine learning technique based on the minimum principle of structural risk and the VC dimension theory of statistical learning theory. It chooses the ideal level of model complexity and learning capacity to optimize promotion ability based on the data properties. In contrast to the conventional linear model, SVM has a simpler calculation procedure which is also appropriate in cases with less data and produces the global optimal search method. Accordingly, SVM can be divided into three types: the One Class SVM for distribution estimation, the Support Vector Regression (SVR) for regression problems, and the Support Vector Classifier (SVC) for classification problems. The following is a brief overview of SVR.

Training samples $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), i = 1,2, \dots n\}, (x_1 \in R^d), (y_i \in R)$, where x_i represents the input column vector corresponding to the ith training sample and y_i is the output value corresponding to x_i. The linear regression function is established as follows:

$$f(x) = w^T \cdot \phi(x) + b \tag{2}$$

where w is the weight vector in the feature space, b is the bias and ϕ is a nonlinear mapping function. The optimization of formula (5) requires a minimum w value, attainable using convex quadratic programming:

$$min\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i - \xi_i^*) \tag{3}$$

These are constraint conditions:

$$s.t.\begin{cases} y_i - w \cdot \phi(x) - b \le \varepsilon + \xi_i \\ w \cdot \phi(x) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i \ge 0 \\ \xi_i^* \ge 0 \end{cases}, i = 1,2,\dots n \tag{4}$$

where the function of C is to maintain a balance between generalization ability and model complexity as penalty parameter, ε indicates the regression function's error requirement, which ensures sparsity and ξ_i, ξ_i^* are two slack variables and restrict the upper and lower bounds of the output value. Their method involves using the Lagrange function and introducing a kernel function for solving the best approach.

By incorporating kernel functions, the nonlinearity of input data issue can be solved. Considering the historical data concerning water consumption incorporates other insights, such as holiday data

and weather information, it is possible to resolve the complicated aspects associated with water consumption in cities by altering the kernel function into an environment with higher dimensions. A nonlinear SVM function model is created by using the following equation.

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x, x') + b \quad (5)$$

where $k(x, x')$ represents the kernel function, and $\alpha_i - \alpha_i^*$ are Lagrange multipliers.

2.6 Different Types of Kernels in SVM

There are four types of kernels that are used in this paper such as Radial Basis Function (RBF) kernel, Sigmoid kernel, Linear kernel, and Polynomial kernel, and where $k(x, x')$ refer to kernel function while x and x' are the input vectors.

2.6.1 RBF kernel

As shown below, the formula of the RBF kernel contains γ as a parameter that determines the spread of the radial basis functions.

$$k(x, x') = e^{(-\gamma\|x-x'\|^2)} \quad (6)$$

2.6.2 Sigmoid kernel

The following expression describes the sigmoid kernel where α and β are user-defined parameters.

$$k(x, x') = tanh(\alpha \cdot x \cdot x' + \beta) \quad (7)$$

2.6.3 Linear kernel

This formula demonstrates how the linear kernel can be defined as the dot product between two input vectors.

$$k(x, x') = x \cdot x' \quad (8)$$

2.6.4 Polynomial kernel

The polynomial kernel is defined as formula below where γ refer to scaling parameter, r is a coefficient and d is the degree of the polynomial.

$$k(x, x') = (\gamma \cdot x \cdot x' + r)^d \quad (9)$$

2.7 Performance Metrics

The following four performance indicators were considered when evaluating the models produced which are Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Coefficient of Correlation (R). In this study, these metrics applied to

check the accuracy and error of the prediction. The equations for evaluation metrics are as the Eqs. (10) to (13) whereas the n is the number of outputs, p_i is the ith predicted output and d_i is the ith observed output.

2.7.1 Mean Absolute Percentage Error (MAPE)

MAPE is used widely to measure accuracy and obtaining significant information from the dataset. The formula of MAPE is shown as below:

$$MAPE = \frac{100}{n}\sum_{i}^{n}\frac{p_i - d_i}{p_i} \tag{10}$$

2.7.2 Mean Absolute Error (MAE)

The following equation is the formula for Mean Absolute Error (MAE) which is usually used in time series forecasting.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - d_i| \tag{11}$$

2.7.3 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is the standard deviation of the residuals where it shows the general overview of the errors that involved in the prediction. Therefore, the lower the RMSE value, the better. The RMSE is determined using the following expression.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - d_i)^2} \tag{12}$$

2.7.4 Coefficient of correlation (R)

In general, the coefficient of correlation (*R*) indicates the degree to which the output seen and expected were related linearly. Increasing similarity between projected and actual outputs increases the correlation coefficient. To achieve better results, the correlation coefficient should be greater. The correlation coefficient (R) can be calculated using the following expression.

$$R = \frac{\sum_{i=1}^{n}\left(p_i - \frac{p_i}{n}\right)\left(d_i - \frac{d_i}{n}\right)}{\sqrt{\sum_{i=1}^{n}\left(p_i - \frac{p_i}{n}\right)^2}\ x\ \sqrt{\sum_{i=1}^{n}\left(d_i - \frac{d_i}{n}\right)^2}} \tag{13}$$

3. Results

An appropriate statistical test should be conducted to prove the observations normality, and the existence of the trend. The Anderson Darling (AD) test checks the normality of the dataset. The *p*-value of the AD test is less than 0.05, as shown in Table 3. Thus, the null hypothesis is rejected which means the data does not follow a normal distribution. Then, the Mann-Kendall test was also used to check the existence of the trend. Since the value of *p* in Mann-Kendall test is smaller than 0.01, the null hypothesis is rejected which indicates that there is trend in the dataset. Next, the dataset

undergoes the pre-processing data where the data is clean and screen. However, the outliers in the dataset are kept because it indicates the extreme values of hydrological data.

Table 3
Statistical tests

Statistical tests	Value	*p*-value
Anderson Darling (AD) test	825.907	<0.05
Mann-Kendall test	-	<0.01

Once the dataset has been pre-processed, it can be used for prediction and forecasting. Table 4 presents the results of the study based on the performance measures used, which are the MAPE, MAE, RMSE and *R* values. During the training set, ANN with 15 hidden nodes exhibits superior performance as it has the lowest MAPE, MAE, RMSE and highest *R* values. The training set is used to train the model, which enables the model to learn patterns, relationships, and features based on the data. The testing set, on the other hand, is used to evaluate the model's performance and to determine its generalization ability to new and previously unknown data. Based on the results of the testing set, it can be observed that the ANN network architecture with 14 nodes shows the lowest MAE, RMSE, and highest R values in comparison with other network architectures. Even though the results of the training period indicated that the model with 2*n*+1 was the most accurate, the network architecture with *2n* hidden nodes was chosen because of its stability.

Next, Table 4 also presents experimental results for SVM. The results from the training and testing sets showed that SVM with a polynomial kernel archived better results than SVM with another kernel. This is primarily because the polynomial kernel is relatively simpler than either RBF or sigmoid kernels. Moreover, polynomial kernels allow the complexity of decision boundaries to be controlled and minimizes the risk of overfitting and underfitting. A polynomial kernel is also capable of capturing the complex relationships between variables without the need to add new features to capture them, which is a weakness of other kernels. RFB kernels also provide competitive results to sigmoid and linear kernels due to their close proximity to each other.

Table 4
Forecast accuracy of the ANN and SVM model during the training and testing period

Model	Number of hidden Nodes / different types of kernels	Training				Testing			
		MAPE	MAE	RMSE	R	MAPE	MAE	RMSE	R
ANN	*n/2*	3.0805	8.9234	15.0004	0.582	2.3422	8.7603	14.2080	0.597
	n+3	4.6899	8.8772	14.9418	0.586	2.3911	8.8076	14.2760	0.592
	2n	6.0954	8.8316	14.9130	0.589	2.4429	8.6899*	14.1347*	0.603*
	2n+1	2.5210*	8.8158*	14.8321*	0.595*	1.6299*	8.9348	14.4611	0.583
SVM	RBF	8.8377	9.0118	18.6280	0.515	1.4678*	9.2126	18.0407	0.533
	Sigmoid	1.4136*	9.2797	19.0746	0.474	1.7608	9.5152	18.5106	0.495
	Linear	2.7837	9.1880	18.6656	0.486	2.8889	9.3768	18.0681	0.513
	Polynomial	2.3981	8.2328*	16.0016*	0.558*	2.0911	8.2225*	15.4304*	0.560*

* Represents the best result among others selection of hidden layers and different kernels

The hydrograph for both the best performing model using ANN and SVM can be seen in Figures 2 and 3. The Figures 2 and 3 illustrate distinct patterns that can explain and describe the degree of fit between the forecast value model and the actual values for both ANN and SVM models. Figure 3 illustrates that some points are under-forecast in comparison with Figure 2. Despite this, both models

shown can follow the fluctuation trend of the daily amount of rainfall throughout the years, despite being not very close to the actual amount of rainfall.

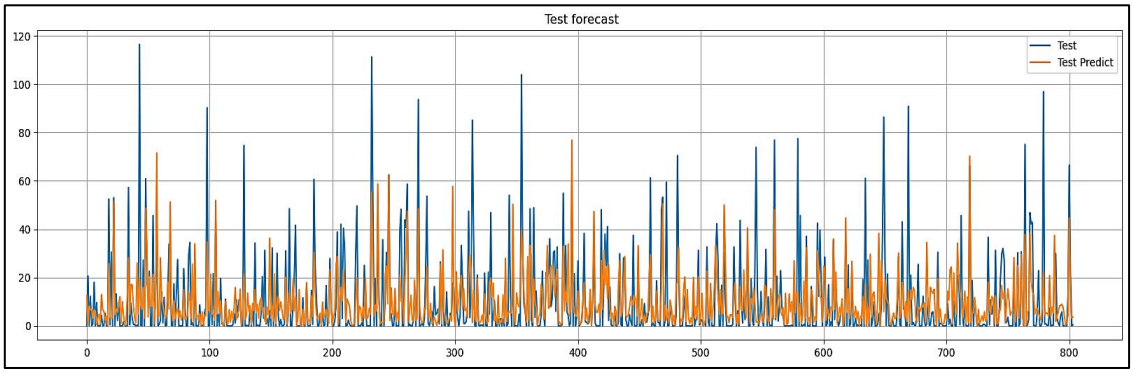

Fig. 2. Observed versus predicted values using ANN model

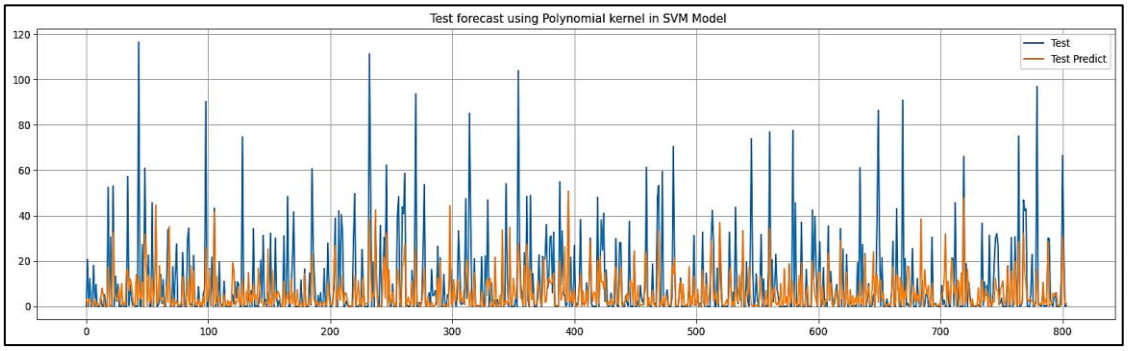

Fig. 3. Observed versus predicted values using SVM model

4. Conclusions

This study tested the hydrological dataset using two statistical tests known as Mann-Kendall and Anderson-Darling tests. Furthermore, the capability of two machine learning models, ANN and SVM, in predicting and analyzing hydrological extreme values was also investigated. The hydrological data used in this study met the assumptions of non-normality and had an existing trend. When analyzing hydrological data, underlying trends are considered based on this assumption. Without considering the trends, it is impossible to accurately predict and analyze extreme values existed in the dataset. It is also possible to determine the appropriate model to be used for prediction purposes based on the results of the statistical tests. Experimental results indicated that ANN performs better than SVM in terms of performance. It is likely that this is because ANN has the capacity to handle complex and non-linear data and are more suited to large-scale datasets than SVM. Moreover, the study indicates that the number of hidden nodes has an impact on performance measurement. Model accuracy can be improved by adjusting the number of hidden nodes in a model, since hidden nodes are a parameter that can be adjusted. It is generally believed that the more hidden nodes the model contains, the more complex it is and the better it is at analyzing large datasets.

Moreover, it was found that SVM with a polynomial kernel performed the best compared to other kernels. It may be attributed to the robustness of the polynomial kernel to outliers, which enables it to capture trends in data and handle complex non-linear interactions more effectively than other kernels. Due to its ability to handle complex non-linear interactions and its robustness to outliers, polynomial kernels are better suited to hydrological data. The findings of this study should be further studied and validated through further experimental analysis.

Acknowledgement
This research was supported by Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2022/ICT06/UTHM/03/1).

References
[1] Leta, Olkeba Tolessa, Aly I. El-Kadi, and Henrietta Dulai. "Impact of climate change on daily streamflow and its extreme values in pacific island watersheds." *Sustainability* 10, no. 6 (2018): 2057. https://doi.org/10.3390/su10062057
[2] Shaharudin, S. M., N. Ahmad, and N. H. Zainuddin. "Modified singular spectrum analysis in identifying rainfall trend over Peninsular Malaysia." *Indonesian Journal of Electrical Engineering and Computer Science* 15, no. 1 (2019): 283-293. https://doi.org/10.11591/ijeecs.v15.i1.pp283-293
[3] Kundzewicz, Zbigniew W. "Climate change impacts on the hydrological cycle." *Ecohydrology & Hydrobiology* 8, no. 2-4 (2008): 195-203. https://doi.org/10.2478/v10104-009-0015-y
[4] Towler, Erin, Dagmar Llewellyn, Andreas Prein, and Eric Gilleland. "Extreme-value analysis for the characterization of extremes in water resources: A generalized workflow and case study on New Mexico monsoon precipitation." *Weather and Climate Extremes* 29 (2020): 100260. https://doi.org/10.1016/j.wace.2020.100260
[5] Dawley, Shawn, Yong Zhang, Xiaoting Liu, Peng Jiang, Geoffrey R. Tick, HongGuang Sun, Chunmiao Zheng, and Li Chen. "Statistical analysis of extreme events in precipitation, stream discharge, and groundwater head fluctuation: Distribution, memory, and correlation." *Water* 11, no. 4 (2019): 707. https://doi.org/10.3390/w11040707
[6] Kalteh, Aman Mohammad. "Enhanced monthly precipitation forecasting using artificial neural network and singular spectrum analysis conjunction models." *INAE Letters* 2 (2017): 73-81. https://doi.org/10.1007/s41403-017-0025-9
[7] Zounemat-Kermani, Mohammad, Okke Batelaan, Marzieh Fadaee, and Reinhard Hinkelmann. "Ensemble machine learning paradigms in hydrology: A review." *Journal of Hydrology* 598 (2021): 126266. https://doi.org/10.1016/j.jhydrol.2021.126266
[8] Zakaria, Muhamad Nur Adli, Ali Najah Ahmed, Marlinda Abdul Malek, Ahmed H. Birima, Md Munir Hayet Khan, Mohsen Sherif, and Ahmed Elshafie. "Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia." *Heliyon* 9, no. 7 (2023). https://doi.org/10.1016/j.heliyon.2023.e17689
[9] Ahmed, Ali Najah, Faridah Binti Othman, Haitham Abdulmohsin Afan, Rusul Khaleel Ibrahim, Chow Ming Fai, Md Shabbir Hossain, Mohammad Ehteram, and Ahmed Elshafie. "Machine learning methods for better water quality prediction." *Journal of Hydrology* 578 (2019): 124084. https://doi.org/10.1016/j.jhydrol.2019.124084
[10] Liu, Xingpo, Yiqing Zhang, and Qichen Zhang. "Comparison of EEMD-ARIMA, EEMD-BP and EEMD-SVM algorithms for predicting the hourly urban water consumption." *Journal of Hydroinformatics* 24, no. 3 (2022): 535-558. https://doi.org/10.2166/hydro.2022.146
[11] Aquil, Mohammad Amimul Ihsan, and Wan Hussain Wan Ishak. "Comparison of machine learning models in forecasting reservoir water level." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 31, no. 3 (2023): 137-144. https://doi.org/10.37934/araset.31.3.137144
[12] Chen, Guoqiang, Tianyu Long, Jiangong Xiong, and Yun Bai. "Multiple random forests modelling for urban water consumption forecasting." *Water Resources Management* 31 (2017): 4715-4729. https://doi.org/10.1007/s11269-017-1774-7
[13] Safiah Yusmah, M. Y., Louise J. Bracken, Z. Sahdan, H. Norhaslina, M. D. Melasutra, Amirhosein Ghaffarianhoseini, S. Sumiliana, and A. S. Shereen Farisha. "Understanding urban flood vulnerability and resilience: A case study of Kuantan, Pahang, Malaysia." *Natural Hazards* 101 (2020): 551-571. https://doi.org/10.1007/s11069-020-03885-1
[14] Aquil, Mohammad Amimul Ihsan, and Wan Hussain Wan Ishak. "Comparison of machine learning models in forecasting reservoir water level." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 31, no. 3 (2023): 137-144. https://doi.org/10.37934/araset.31.3.137144
[15] Govindaraju, Rao S. "Artificial neural networks in hydrology. II: hydrologic applications." *Journal of Hydrologic Engineering* 5, no. 2 (2000): 124-137. https://doi.org/10.1007/978-94-015-9341-0
[16] Aydın, Metin Mutlu, Mehmet Sinan Yıldırım, Orhan Karpuz, and Kiarash Ghasemlou. "Modeling of driver lane choice behavior with artificial neural networks (ANN) and linear regression (LR) analysis on deformed roads." *Computer Science* 4, no. 1 (2014): 47-57. https://doi.org/10.5121/cseij.2014.4105
[17] Tanty, Rakesh, and Tanweer S. Desmukh. "Application of artificial neural network in hydrology—A review." *Int. J. Eng. Technol. Res* 4, no. 6 (2015): 184-188.
[18] Dolling, Oscar R., and Eduardo A. Varas. "Artificial neural networks for streamflow prediction." *Journal of hydraulic research* 40, no. 5 (2002): 547-554. https://doi.org/10.1080/00221680209499899

[19] Kuok, Kuok King. "Parameter optimization methods for calibrating tank model and neural network model for rainfall-runoff modeling." PhD diss., Ph. D. Thesis. University Technology Malaysia, 2010.
[20] Banhatti, Aniruddha Gopal, and Paresh Chandra Deka. "Effects of data pre-processing on the prediction accuracy of artificial neural network model in hydrological time series." *Urban Hydrology, Watershed Management and Socio-Economic Aspects* (2016): 265-275. https://doi.org/10.1007/978-3-319-40195-9_21
[21] Araghinejad, Shahab, Mohammad Azmi, and Majid Kholghi. "Application of artificial neural network ensembles in probabilistic hydrological forecasting." *Journal of Hydrology* 407, no. 1-4 (2011): 94-104. https://doi.org/10.1016/j.jhydrol.2011.07.011
[22] Mo, Chongxun, Zhiwei Yan, Rongyong Ma, Xingbi Lei, Yun Deng, Shufeng Lai, Keke Huang, and Xixi Mo. "Investigation of the EWT–PSO–SVM Model for Runoff Forecasting in the Karst Area." *Applied Sciences* 13, no. 9 (2023): 5693. https://doi.org/10.3390/app13095693
[23] Han, D., L. Chan, and N. Zhu. "Flood forecasting using support vector machines." *Journal of hydroinformatics* 9, no. 4 (2007): 267-276. https://doi.org/10.2166/hydro.2007.027
[24] Hirani, Dhawal, and Nitin Mishra. "A survey on rainfall prediction techniques." *International Journal of Computer Application* 6, no. 2 (2016): 28-42.
[25] Zhang, Jinlei, Xue Qiu, Xiang Li, Zhijie Huang, Mingqiu Wu, and Yumin Dong. "Support vector machine weather prediction technology based on the improved quantum optimization algorithm." *Computational Intelligence and Neuroscience* 2021, no. 1 (2021): 6653659. https://doi.org/10.1155/2021/6653659