

FEEMD AND GWO METHODOLOGY FOR FLOOD EARLY WARNING PREDICTION MODEL

NOOR HAYATI MOHD ZAIN¹, NORAFIDA ITHNIN²

^{1,2}School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor, Malaysia

E-mail: ¹noorhayati5687@yahoo.com, ²afida@utm.my

ABSTRACT

Flood disasters are natural hazards that cause many great losses either in terms of lives, property, and even the structure of the earth's surface. Investigations on this topic have become one of the ongoing studies because it has a great impact on the environment and community life. This study also highlights the improvement of flood warning measurement methods to ensure that the adverse effects of flood disasters can be controlled better than before. This paper will present a more efficient method to model flood early warning prediction in Malaysian districts particularly. This study focuses on the use of Fast Ensemble Empirical Mode Decomposition (FEEMD) to decompose selected rainfall dataset. Furthermore, the Gray Wolf Optimizer (GWO) was used as an optimization approach to optimize between FEEMD hybrids with Artificial Intelligence models to find the most accurate flood early warning prediction model. The study also aims to improve the process flow of the flood early warning prediction system delivered to flood victims by ensuring that they are able to prepare for the consequences of impending flood disasters in their areas of residence.

Keywords: *Flood Warning Prediction, Natural Disaster, Monsoon Flood, FEEMD, GWO*

1. INTRODUCTION

Flood refers to a geographical area of land that is covered by water and submerges the surrounding areas that are not normally covered by water (Shahabi *et al.*, 2020). The increase in flood disasters has occurred around the world due to various causes either from human hands or from changes in nature. Globally, the heavy rains of the northeast monsoon also led to flash floods and landslides in small geographical catchments usually around 20–30 km² that damaged the areas affected by the floods (Venkatappa *et al.*, 2021; Vijaykumar *et al.*, 2021). This proves that research in this field is one of the topics that is in line with international studies and can also be used as a reference in the future by globally.

In Malaysia, the local climate is governed by two main elements, namely the southwest and northeast monsoons. The southwest monsoon refers to the heavy rains that fall between May and August. While the northeast monsoon refers to heavy rains that fall starting in November and ending in February. The northeast monsoon that causes the

monsoon floods that usually hit the east coast of Peninsular Malaysia (Mohamad *et al.*, 2020).

In Malaysia, the Department of Irrigation and Drainage (DID) is one authorized department that responsible for taking the role of in-charge for collecting water level data and rainfall data for doing flood prediction for flood warning system. Even though, there is still a gap because there is no systematic approach exists for doing prediction to be used by DID (Esa, 2018). This shown that Malaysia needs to improve delivery warning system to avoid a huge flood damage in the future because of changing climate keep happening with a different pattern.

This study is suggested to focus on the flood early warning prediction model by using the hybridization of statistic and soft computing methods in estimating the prediction accuracy error with inclusion of time series rainfall dataset in Malaysia which choose Johor as the data collection area of study. To be more specific, the algorithm by GWO been used as the optimizer for the suggested hybrid method of this study which are adaptive network-based fuzzy inference system (ANFIS), Artificial neural networks (ANN), and fuzzy rule-

based (FRB). Then, this study used daily rainfall dataset from Department of Irrigation and Drainage, Nusajaya, Johor. Figure 1 is presenting the figure of Johor state.



Figure 1: The state of Johor

The objective of this study is to identify the most accurate method to be used as a model of flood early warning prediction model particularly in Malaysia. The experimental analysis used involved time series dataset that were entered into the ANN algorithm as reference techniques before being adapted to the GWO, FEEMD, and ANFIS techniques. These techniques are combined and analyzed for each data result that has been processed according to the most accurate accuracy values. This builds on a better flood early warning prediction model in Malaysia than the existing one. This paper is highlighted the three phases of this study framework flow in order to achieve the objective in identifying the more accurate method for early warning prediction model.

2. PROBLEM BACKGROUND

Natural hazards such as flood is not an easy thing to predict. Great hazards are also known as disaster and it is brought a great danger. Through newspapers and the internet, many stories mention that most countries around the world will not able to escape facing disaster. There are various types of disasters that occur around the world every year such as floods, droughts, tsunamis, hurricanes, and even forest fires. Malaysia also not left behind in facing disasters. The most frequent disaster in Malaysia is the flood

disaster (Abdul and Mohamed, 2013; Mehedi et al., 2018; Y et al., 2018; Shahid Latif, 2020). Many measurement methods have been taken to reduce the impact of these floods (Rasmi and Puad, 2014; Omar and Chong, 2018; Sadeka et al., 2020). These studies prove that there is a precise reference relationship between the topic of early warning of floods with the structure from previous studies that highlight the control of the effects of flood disasters. However, it is remain in unsatisfactory solutions (Rasmi and Puad, 2014; Yen et al., 2020).

Many research on flood disasters has been done to determine how to reduce or prevent them from events that can cause adverse effects on people and the environment (Abdul and Mohamed, 2013; Rasmi and Puad, 2014; Mehedi et al., 2018; Omar and Chong, 2018; Y et al., 2018; Chawla and Mujumdar, 2020; Kumar et al., 2020; Shahid Latif, 2020). As shown in Malaysia, there are many research studies conducted related to flood disaster (Abdul and Mohamed, 2013; Rasmi and Puad, 2014; Mehedi et al., 2018; Omar and Chong, 2018; Y et al., 2018; Sadeka et al., 2020; Shahid Latif, 2020; Yen et al., 2020; Brito et al., 2021). It is included the study regarding flood early warning prediction model. However, the flood early warning prediction model remains open for any improvements. This is because the flood early warning prediction model is still depending on an ambiguous prediction system.

3. RESEARCH METHODOLOGY FRAMEWORK

The methodology of this study involves three phases of the process. The first phase refers to the phase to conduct investigation and clarification process for daily rainfall dataset that has been taken from DID Johor. Daily rainfall dataset is seen to be more relevant in this study for the prediction of time series. Related to this, the general architecture of the GWO, FEEMD, and ANFIS algorithms has been described as an appropriate approach in developing a flood early warning prediction model for studies using daily rainfall dataset in the second phase.

Then, to propose and plan a more relevant flood early warning prediction model in Malaysia, the third phase is carried out which involves the process of conducting experiments that have been designed in the previous phase and the evaluation process followed after that. Figure 2 shows the three phases of the research framework involved in this study.

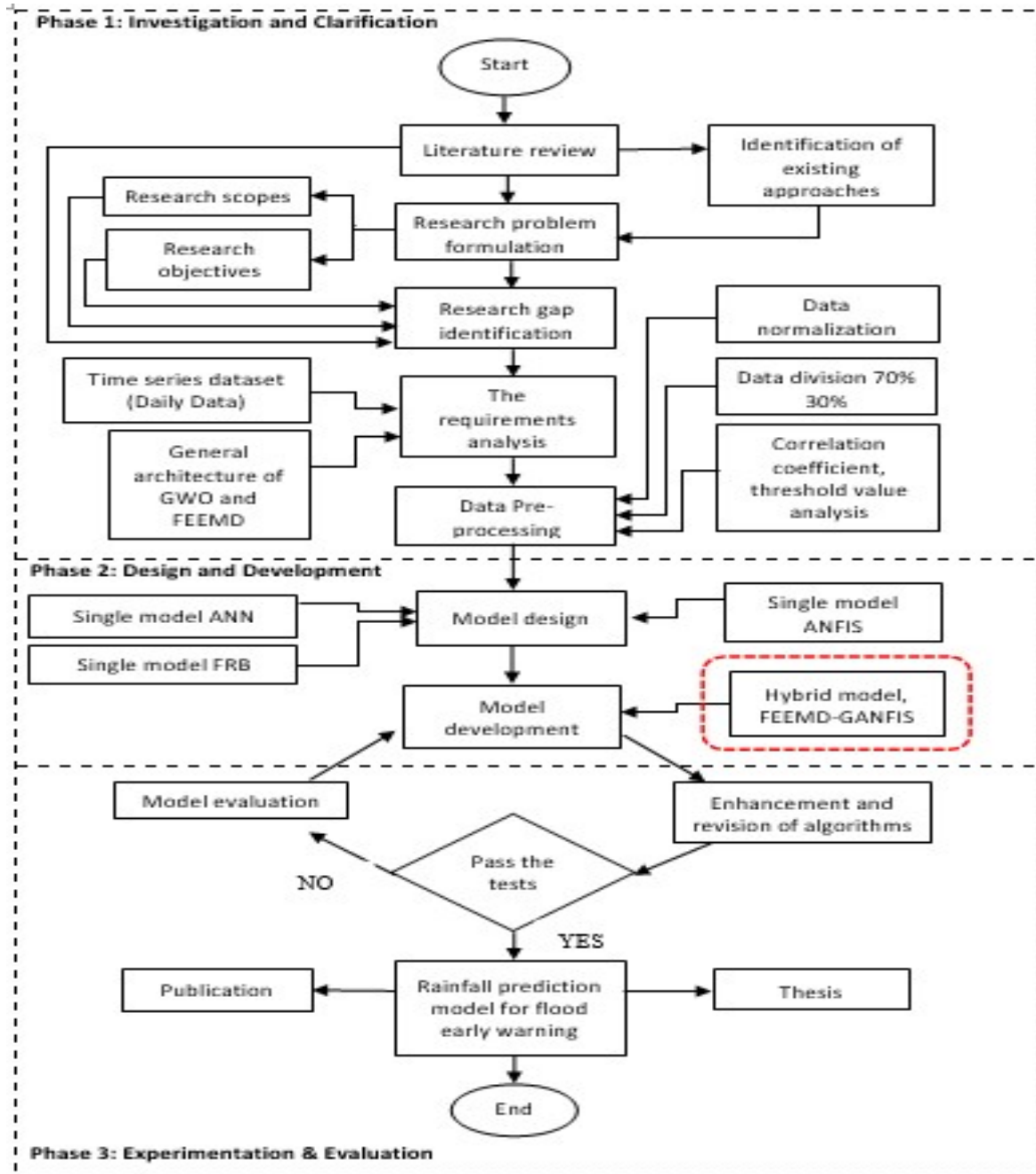


Figure 2: The Research Framework

Based on the research framework, the previous studies of ANN model and FRB model are set as guidance to develop new flood early warning prediction model for this research. As related, the propose flood early warning prediction model need to obtain approval in research and development

phase. The model then first was tested with unit testing to test the detection of flood early warning prediction. If some errors found further enhancement will be made on the model to make sure the model designed achieve all research objectives. Moreover, each phases of framework

already been design to control the research flow in achieving the objectives. The following sub-sections are explaining how each phases handling the research objectives.

3.1 Phase 1: Investigation and Clarification

In this phase, the existing academic research materials where been explored by the literature review technique. Then it is follows by research problem formulation, research gap identification, and end up this phase with initialized the study requirement analysis. In addition, the data pre-processing was also conducted in the research Phase 1 to clarify suitable dataset for prediction model training and testing. Overall, this Phase 1 is for achieving the objective for “investigate the existing approaches, formulate research problem, and identify research gap.” Table 1 is presenting the list of activities for the Phase 1.

Table 1: Phase 1’s Activities

Activities		Deliverable
Literature review	Investigate the early warning prediction models	Early warning prediction models identified
	Investigate the prediction methods for flood early warning prediction model	The prediction methods for flood early warning prediction model identified
	Investigate the approaches to quantify the accuracy of the flood early warning prediction model	The approaches to quantify the accuracy of the flood early warning prediction model identified
	Investigate the issues presents in flood early warning prediction	The issues present in flood early warning prediction identified
Research problem formulation	Clarifies research problem, objectives and scopes	The formulation of research scopes, problem, and research objectives
Research gap	Clarifies the	The General

identification	research gap	architecture of GWO and FEEMD have been overlooked
Requirements analysis	Clarifies time series dataset of General architecture of GWO and FEEMD	The time series dataset of General architecture of GWO and FEEMD clarified
Data pre-processing	Select the Correlation coefficient, threshold value analysis.	Correlation coefficient, threshold value analysis selected
	Clarifies Data division 70% 30%	The Data division 70% 30% clarified
	Clarifies Data normalization	The data normalization clarified

3.2 Phase 2: Design and development

This Phase 2 of design and development consists of two main research activities, which are the model design and model development. In this research phase, the ANN model, FRB model, and ANFIS model were reviewed from previous studies to be used as guideline for model design before develop the flood early warning prediction model in this research. Overall, this Phase 2 is for achieving the objectives for “designing and develop the research algorithm” and for “designing and develop the theoretical model”. Table 2 is presenting all the activities which be done through this Phase 2.

Table 2: Phase 2’s Activities

Activities		Deliverable
Model and algorithm design & development	Design single model ANN, FRB, ANFIS, Hybrid model, FEEMD-GANFIS.	The model and algorithm developed.
	Develops the model and algorithm for flood early warning prediction.	A model of flood early warning prediction with more accurate level.

3.3 Phase 3: Experimentation and Evaluation

In research phase of experimentation and evaluation, the proposed model of flood early warning prediction in this research were validated. The proposed model had been benchmarked in this research activity as a model of flood early warning prediction. After the model development, the model was tested with unit testing to test the detection of flood early warning prediction. If there are some errors found in the model developed, further enhancement will be made on the model to make sure the model designed achieve the research objectives. For each verified model, it was integrated into the existing testing frameworks to design the model of flood early warning prediction. Then, the developed model of flood early warning prediction was validated. The model requirement analysis framework is presented in Figure 3.

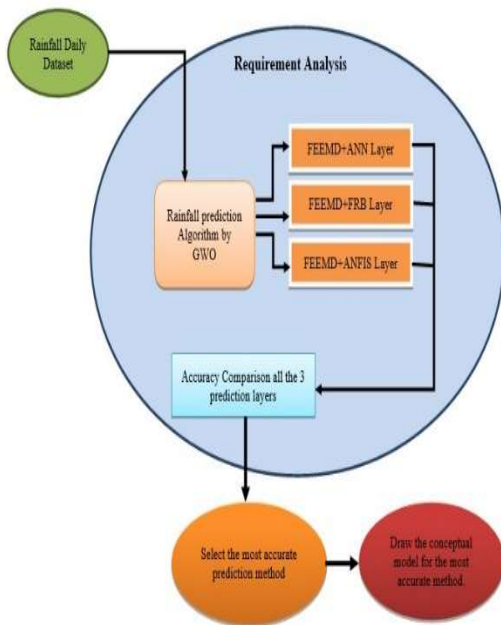


Figure 3: Requirement Analysis Framework

This testing framework consists of four main steps, which includes preparation, execution, verification and data analysis. The preparation step is to select existing ANN model, FRB model, and ANFIS model before develop the model in this research work. Next, second step which is refer to execution, it involves the selected models that reviewed from previous literature to configured what variables can be used in the model

developed. Then, execution step was performed to conduct manual testing to clarify the validation of test results. Last phase, the data analysis was performed with collected dataset were gathered, filed and recorded. Overall, this Phase 3 is for achieving the objectives for “*benchmarking the model accuracy*”. Table 3 is presenting all the activities which be done through this Phase 3.

Table 3: Phase 3’s Activities

Activities	Deliverable	
Enhancement and revision of algorithms	Revises the algorithms accuracy	Evaluation of flood early warning prediction model
	Enhances and revises error codes and bugs	The model that meets the desire specifications
Model evaluation	Evaluate the models developed	The models are evaluated
Thesis (Research finding documentation)	Thesis writing	A copy of thesis
Publication	Academics paper and journal writing	The academic research paper and journal

4. TYPES OF RESEARCH: EXPERIMENTAL RESEARCH

In order to provide a clear guidance in conducting this research; a framework that contains research design is built in order to conduct the full process of this research. The research issues are solved by research methodology that comprises of,

- i. Investigations is been carried out to seek for clarification on the issues
- ii. Conducting a research and development of latest model design
- iii. Performing evaluation process for quality assessment

5. SIGNIFICANCE OF RESEARCH

As familiarly in research field, there are many studies which going through the details about flood early warning system. Even though, this research takes a good approach in proposing

the better method by implemented the hybrid method of FEEMD-GANN, FEEMD-GFRB, and FEEMD-GANFIS comparison approach for flood early warning prediction model in Malaysia. The situation where none of the previous related study which used this kind of method give the great opportunity to this study. Benefits from this, more life and fast preparation can be done before facing flood disasters in future by all related agencies includes the public community.

Flood early warning prediction model helps in assisting people to raise awareness and be prepare on any consequences focus on high-risk area. Hence, the following are the targeted contribution from this research study improve the flood early warning prediction model in Malaysia.

6. RESEARCH THEORIES

In completing this study, several related theories been referred. The following sub-sections is presenting all the related theories which been used in completing this study and achieving the main target of this study to provide the accurate flood early warning prediction model.

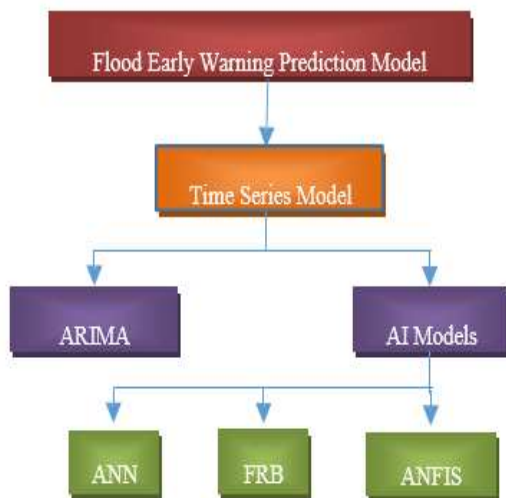


Figure 4: Theories Review

6.1 ARIMA Time Series Models

The time series model is significant and useful to model data that are collected and indexed by time. It has the best fit to each of the dataset is selected. Time series model can be used to assess the trend of rainfall characteristics and conditions of climate parameters. Time series model used for analysis of time series in scientific applications.

Time series modelling method is suitable and essential for prediction because it enables instant correlation between multiple factors. Time series model is the method that can widely applied in the study and it is scalable. Time series model is mostly used because it is a straightforward calculations and good approximation long history of successful applications.

ARIMA is stand for Auto Regressive Integrated Moving Average Model. Basically, ARIMA model is based on ARMA model on different time series data where the number of times differencing is performed on the original time series data is denoted by 'd' to make the data stationary. To find out value of p and q, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) are applied on the differenced data. If both ACF and PACF have sinusoidal decay, and become zero after lags q and p respectively, then process order for ARMA model is (p,q) and process order of ARIMA model is (p,d,q) (Babu and Reddy, 2012). ARIMA model is used to predict non-stationary time series. By utilizing ARIMA model to studying time series, if the time series is non-stationary, the difference for the series is done because it is transformed into stationary time series. ARIMA model is in the process of modelling that does not depend on the series value, but it is depend on the dynamic statistical characteristics between the sequence values (Yang et al., 2013).

6.2 Artificial Intelligence (AI) Models

Artificial intelligence (AI) models can be use in natural disaster to improve flood residence and preparedness. The benefit for use in natural disaster is can provide suitable tools for natural disaster prediction and impact assessment where it can help to provide better preparedness and prevention for face natural disaster. AI can be used during flood events for emergency responders to have situational awareness (Saravi et al., 2019). Besides, AI model is good for prediction accuracy (Afan et al., 2020).

AI methods are nonparametric models in which inputs are mapped into outputs and the underlying process involved is not explored. AI models have the ability to learn complex and non-linear relationship which is difficult for conventional time series models to learn. Our reviews on this shall cover three types of AI models which are Artificial Neural Network (ANN), Fuzzy Rule Based (FRB) and Adaptive Neuro Fuzzy Inference System (ANFIS) models.

6.3 Artificial Neural Network (ANN) Models

Artificial Neural Network (ANN) is good candidates for quantitative prediction, which is concerned with extracting patterns from past events and extrapolating them into the future. The branch of artificial intelligence that considers classification as one of the most dynamic research and application areas. Development of artificial neural network began with motivated by a desire to understand brain and to emulate some of its strengths. Based on their ability of learning, training, simulation and prediction of data. Neural network has several types of layer, which is input layer, hidden layer and output layer.

ANN has the ability of capturing the autocorrelation structure of the time series even if the underlying law is known or too complex to describe. ANN is good candidates for quantitative prediction, which is concerned with extracting patterns from past events and extrapolating them into the future. Some common ANN models are consisting of multilayer feed-forward neural network (MFFNN), radial basis function neural network (RBFNN) and recurrent neural network (RNN). Artificial neural network (ANN) is the branch of artificial intelligence that considers classification as one of the most dynamic research and application areas. The development of artificial neural network began with motivated by a desire to understand brain and to emulate some of its strengths. An artificial neural network is a processing system of information that has certain perform features with biological neural networks (Al-Sammarraie et al., 2018). Artificial neural networks (ANN) are identified based on their ability of learning, training, simulation and prediction of data. Neural network has several types of layer, which is input layer, hidden layer and output layer. A computational neural network consists of simple processing units called as neurons. Each of neurons in neural network receives information corresponding to one of the independent variables used as input.

6.4 Fuzzy Rules Based (FRB) Models

Fuzzy rule-based (FRB) modelling incorporates all possible scenarios of input parameters. In fuzzy set theory, an element has the membership grade between 0 and 1, and while in a traditional set, the membership grade is either in 1 or 0. The fuzzy rule based modelling follows the logic “if- then” where “if” is used for antecedents and “then” is used for consequences (Chowdhury

and Fahmi, 2018). According to previous scholar, the early warning system for flood disaster based on fuzzy logic model in IoT system. The fuzzy logic model consists of input and output. The fuzzy model inputs are the rainfall and height of water. Then, the output of the fuzzy model is the status of flood disaster (Yuwono et al., 2020). In crisp logic, the premise x is A can only be true or false. However, in a fuzzy rule, the premise x is A and the consequent y is B can be true to a degree, instead of entirely true or entirely false. This is achieved by representing the linguistic variables A and B using fuzzy sets.

6.5 Adaptive Neuro Fuzzy Inference System (ANFIS) Models

Adaptive Neuro Fuzzy Inference System (ANFIS) is one of the computational methods which are introduced by Lotfi A. ANFIS model is using Takagi-Sugeno fuzzy inference system. This method has a hybrid algorithm to estimate the parameters, least square to estimate the linear parameter and back propagation to estimate nonlinear parameter. Suitable for prediction using ANFIS for predict daily rainfall to be more accurate and got significant results. The ANFIS model performed using the ‘subtractive fuzzy clustering’ functions to perform successfully.

7. GEOGRAPHICAL RESEARCH AREA

The study decided to focus on the areas that received the highest rainfall distribution. Daily rainfall data obtained from the DID Johor has shown that there are three areas in Johor receive highest amount of rainfall and are involved with floods during the southwest monsoon period in Malaysia. The Figure 5 is presenting the chart of the rainfall distribution obtained.

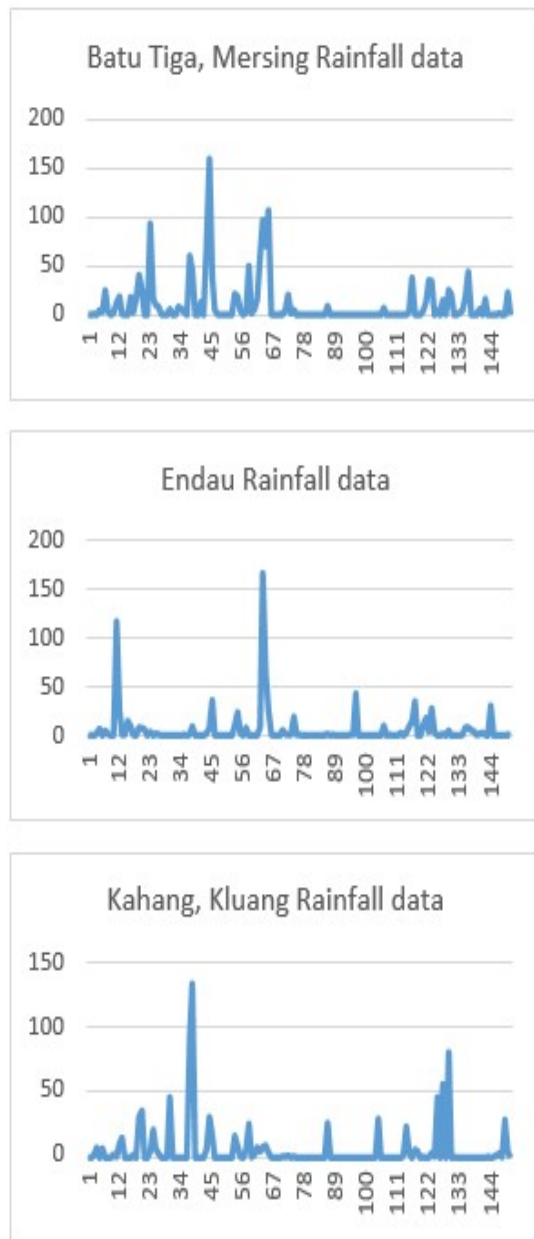


Figure 5: The Selected of Rainfall Data Distribution

As shown in Figure 5, the areas involved in this study are Endau, Batu 3 Mersing, and Kahang Kluang from Johor region. Meanwhile, Figure 6 shows a map of the three areas.

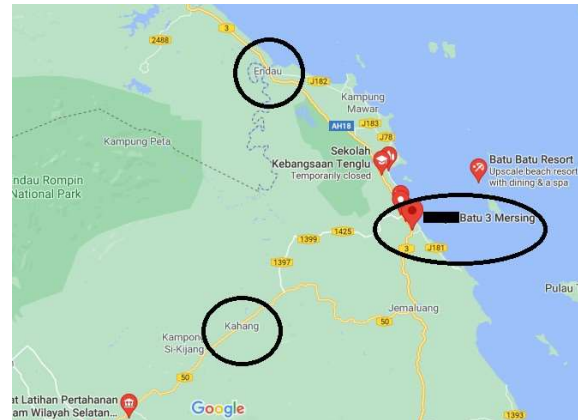


Figure 6: The Selected Flood Area

8. CONCLUSION

This paper successfully highlights the three flow phases of the framework as the major finding through this paper in achieving the objective of identifying a more accurate method for early warning flood prediction model. Future work to continue this paper is to continue the construction of a prototype for a flood early warning prediction model that follows the hybrid method described in this paper by using rainfall data that occurs in Malaysia. This is significantly a much-needed study for flood victims' residential areas so that they are better prepared when the flood season arrives based on a more accurate warning system than before. This study will also fill the gaps in the Malaysian region that have been identified previously because of the early flood prediction conditions that need to be improved to reduce the danger of flood disasters that cause great losses to local communities and the country in particular.

REFERENCES

- [1] Abdul, M. and Mohamed, G., "Mitigation of Climate Change Effects through Non-structural Flood Disaster Management in Pekan Town, Malaysia", *Procedia - Social and Behavioral Sciences*. Elsevier B.V., 85, 2013, pp. 564–573.
- [2] Afan, H. A., Allawi, M. F., El-Shafie, Amr, Yaseen, Z. M., Ahmed, A. N., Malek, M. A., Koting, S. B., Salih, S. Q., Mohtar, W. H. M. W., Lai, S. H., Sefelnasr, A., Sherif, M. and El-Shafie, Ahmed, "Input attributes optimization using the feasibility of genetic nature

- inspired algorithm: Application of river flow forecasting”, *Scientific Reports*, 10(1), 2020, pp. 1–15.
- [3] Al-Sammarraie, N. A., Al-Mayali, Y. M. H. and Baker El- Ebiary, Y. A., “Classification and diagnosis using back propagation Artificial Neural Networks (ANN) algorithm”, *2018 International Conference on Smart Computing and Electronic Enterprise, ICSCEE 2018*, 2018, pp. 10–14.
- [4] Babu, N. and Reddy, B. E., “Temperature Using Variants of ARIMA models”, *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012)*, (1), 2012, pp. 256–260.
- [5] Brito, D., Marques, R., Augusto, C., Moura, R., Neto, B., Hugo, V. and Coelho, R., “Monitoring meteorological drought in a semiarid region using two long-term satellite-estimated rainfall datasets : A case study of the Piranhas River basin , northeastern Brazil”, 250 (July 2020).
- [6] Chawla, I. and Mujumdar, P. P., “Evaluating rainfall datasets to reconstruct floods in data-sparse Himalayan region”, *Journal of Hydrology*. Elsevier, 588(November 2019), 2020, p. 125090.
- [7] Chowdhury, S. and Fahmi, M. I., “Fuzzy Rule-Based Assessment of Runoff for Abha, Saudi Arabia, Using the Watershed Modeling System (WMS) Software”, *Arabian Journal for Science and Engineering*. Springer Berlin Heidelberg, 43(10), 2018, pp. 5457–5468.
- [8] Esa, J., “Temubual Pengurusan Bencana”, *Pegawai Hidrologi Negeri, Bahagian Sumber Air dn Hidrologi*, (20), 2018, pp. 2–4.
- [9] Kumar, K. C. A., Reddy, G. P. O. and Masilamani, P., “A tool to monitor agricultural drought by using time-series datasets of space-based earth observation satellites”, *Advances in Space Research*. COSPAR, 2020.
- [10] Mehedi, M., Sackor, A. S., Alam, A. S. A. F., Al-amin, A. Q., Bashawir, A. and Ghani, A., “Community responses to flood risk management – An empirical Investigation of the Marine Protected Areas (MPAs) in Malaysia”, *Marine Policy*. Elsevier Ltd, 97(September), 2018, pp. 119–126.
- [11] Mohamad, S., Syed, S., Noorani, M. S. and Razak, F. A., “An Early Warning System for Flood Detection Using Critical Slowing Down”, 2020.
- [12] Omar, N. and Chong, O., “Framework Considerations for for Community Resilient Towards Towards Disaster in Malaysia”, 2018, pp. 1–8.
- [13] Rasmi, A. and Puad, A., “Emergency planning and disaster recovery in Malaysian hospitality industry”, 144, 2014, pp. 45–53.
- [14] Sadeka, S., Suhaimi, M. and Kabir, S. “Disaster experiences and preparedness of the Orang Asli Families in Tasik Chini of Malaysia : A conceptual framework towards building disaster resilient community”, *Progress in Disaster Science*. The Author(s), 6, 2020, p. 100070.
- [15] Saravi, S., Kalawsky, R., Joannou, D., Casado, M. R., Fu, G. and Meng, F., “Use of artificial intelligence to improve resilience and preparedness against adverse flood events”, *Water (Switzerland)*, 11(5), 2019.
- [16] Shahabi, H., Shirzadi, A., Ghaderi, K., Omidvar, E., Al- Ansari, N., Clague, J. J., Geertsema, M., Khosravi, K., Amini, A., Bahrami, S., Rahmati, O., Habibi, K., Mohammadi, A., Nguyen, H., Melesse, A. M., Ahmad, B. Bin and Ahmad, A., “Flood detection and susceptibility mapping using Sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on K-Nearest Neighbor classifier”, *Remote Sensing*, 12(2), 2020, pp. 1–30.
- [17] Shahid Latif, F. M., “Bivariate joint distribution analysis of the flood characteristics under semiparametric copula distribution framework for the Kelantan River basin in Malaysia”, *Journal of Ocean Engineering and Science*. Elsevier B.V, 2020.
- [18] Venkatappa, M., Sasaki, N., Han, P. and Abe, I., “Science of the Total Environment Impacts of droughts and floods on croplands and crop production in Southeast Asia – An application of Google Earth Engine”, *Science of the Total Environment*. Elsevier B.V., 795, 2021, p. 148829.
- [19] Vijaykumar, P., Abhilash, S., Sreenath, A. V, Athira, U. N., Mohanakumar, K., Mapes, B. E., Chakrapani, B., Sahai, A. K., Niyas, T. N. and Sreejith, O. P., “Kerala floods in consecutive years - Its association with mesoscale cloudburst and structural changes in monsoon clouds over the west coast of India”, 33, 2021.
- [20] Y, D. J., S, Y. W., C, N. K. and F, F. J., “Analysis of the Cyclonic Vortex and Evaluation of the Performance of the Radar

- Integrated Nowcasting System (RaINS) during the Heavy Rainfall Episode which Caused Flooding in Penang , Malaysia on 5 November 2017”, 2018, pp. 217–229.
- [21] Yang, R., Li, L., Zhao, Z. and Zhang, Y., “Simulation of rain attenuation time series by ARIMA model”, *2013 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference, CSQRWC 2013*, (4), 2013, pp. 304–307.
- [22] Yen, H., An, C. and Lim, K., “Disaster relief work: The experiences of volunteers in Malaysia”, 43(133), 2020.
- [23] Yuwono, T., Ruzardi, Nordin, R., Mohamed, R. and Ismail, M., “The Model of Fuzzy Logic in IoT System as Decision Support System for Determining Flood Disaster Status”, *Journal of Physics: Conference Series*, 1529(5), 2020.