Flood disaster and early warning: application of ANFIS for river water level forecasting

Amrul Faruq1, Aminaton Marto2, Nadia Karima Izzaty3, Abidemi Tolulope Kuye4, Shamsul Faisal Mohd Hussein5, Shahrum Shah Abdullah4*

Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia1,2,5,6
Department of Electrical Engineering, Faculty of Engineering, Universitas Muhammadiyah Malang, Malang 65144, Indonesia1
Indonesia Ministry of Public Works and Housing, Jakarta 12110, Indonesia3
UNESCO-IHE Institute for Water Education, Delft, Netherlands4

Article Info
Keywords:
Flood Forecasting, ANFIS, Machine Learning, Disaster Risk Reduction, Early Warning

Abstract
Intensively monitoring river water level and flows in both upstream and downstream catchments are essential for flood forecasting in disaster risk reduction. This paper presents a developed flood river water level forecasting utilizing a hybrid technique called adaptive neuro-fuzzy inference system (ANFIS) model, employed for Kelantan river basin, Kelantan state, Malaysia. The ANFIS model is designed to forecast river water levels at the downstream area in hourly lead times. River water level, rainfall, and river flows were considered as input variables located in upstream stations, and one river water level in the downstream station is chosen as flood forecasting point (FFP) target. Particularly, each of these input-output configurations consists of four stations located in different areas. About twenty-seven data with fifteen minutes basis recorded in January 2013 to March 2015 were used in training and testing the ANFIS network. Data preprocessing is done with feature reduction by principal component analysis and normalization as well. With more attributes in input configurations, the ANFIS model show better result in term of coefficient correlation (R) against artificial neural network (ANN)-based models and support vector machine (SVM) model. In general, it is proven that the presented ANFIS model is a capable machine learning approach for accurate forecasting of river water levels to predict floods for disaster risk reduction and early warning.

1. Introduction
Flood disasters are extensive hydrologic experiences resulting in significant property losses and environmental damages as well as the loss of life across the globe. Flood models and mathematical techniques are increasingly important subjects in the state-of-the-art of hydrologic area, especially to simulate river water level and streamflow to predict floods. Research on the advancement of flood forecasting is on the increase since it contributes to disaster risk reduction, which is difficult, challenging, and complex to model [1]. According to Sendai Frameworks for disaster risk reduction (SFDRR) 2015-2030, the DRR is stated in priority number three and four, which are ‘investing in disaster risk reduction for resilience’ and ‘enhancing disaster risk preparedness for effective response’ respectively [2]. Hence, in connection with these viewpoints, flood modeling and forecasting is crucial for disaster risk management. In many regions in the world, flood forecasting is one of the few feasible options to manage flood disasters.

To date, a number of flood forecasting models are mainly data-specific and involve simplified various input assumptions [3]. Thus to mimic the complex mathematical expression of physical processes and river behaviors, such models benefit from specific techniques e.g., empirical black-box models, stochastic and hybrids [4]. These physically and statistically based models boost the usage of advanced data-driven methods, e.g., machine learning techniques. Among them, the most well-known works of flood forecasting model include artificial neural networks (ANNs) [5][6][7], support vector machines (SVMs) [8][9] and adaptive neuro-fuzzy inference system (ANFIS) [3][10], that were effectively employed for both short-term and long-term flood forecasting.

ANN model provides large flexibility in solving non-linear problems, and it has been successfully applied in various hydrological areas [11][12]. ANN has been used for flood forecasting due to its ability and efficiency in terms of computing time. Although ANNs performed more efficiently for solving time series hydrological data rather than physically based models, some attempts are facing difficulties working with ANN methods when choosing ANN model design, input variable and type of datasets, and fit the features into the networks [13]. Modern ANN approach or...
sometimes called hybrid model such as ANFIS also been particularly effective in improving flood forecasting technique due to its high accuracy and capability [14]. These models are fast and reliable. Further advantage of the method is that it can handle dynamic model, highly non-linear, data disturbances, and complex to understand.

Ghaderi Kamal, et.al [15] has demonstrated that ANFIS could perform for good and close performance for flood modelling and frequency analysis in Iran region. On the other hand, Perera and Lahat has implemented flood forecasting model based on fuzzy logic approach [16]. The fuzzy logic approach as discussed provided satisfactory result for river water forecasting. However, only for the river water level as inputs and predicted river water level addressed to the fuzzy model without any other relevant variables such as rainfall and stream flow. In which, these considerable inputs could reflect the performance of the model. Recently, Yaseen, et.al [17] explored machine learning models used for hourly river flow forecasting. It is presented relatively evidences the ability of the proposed models.

Despite the achievement of ANFIS and further related studies, most studies of flood forecasting have barely carried out in a limited to small number of one river basin, particularly a number of rainfall data [14][18]. Their performance issues employing multivariable inputs need further detailed consideration. It is the intention of this work to investigate the performances of these ANFIS model against other models such as ANN and linear regression models in predicting river water levels for flood forecasting problems. This study is to build multi-time ahead data-driven models with considering multivariable inputs data that enable to simulate and predict river water level from historical-observed data using ANFIS model. The study aimed to expand on the results of previous study [19] in which two machine learning algorithms namely radial basis function and non-linear autoregressive exogenous neural networks have been successfully examined.

2. Research Method
2.1 Hydrological Data and Study Area

The proposed method has been evaluated using a case study in Malaysia namely Kelantan River, as a representative of flood forecasting point (FFP). In fact, this reservoir is among the most frequent occurrence of seasonal-flood disasters in Malaysia. The state of Kelantan belongs to the eastern region and located in the northeast of peninsular Malaysia with Kota Bharu is the capital city of Kelantan. Kelantan state facing the bounded with China South Sea in the northeast, while in the east with Terengganu state, while Pahang, Perak, in the south and west respectively, and bounded with Thailand in the north. Kelantan state has a total area about 15,101 km$^2$ and the population is about 1.76 million in year 2015 [20].

© 2021 The Authors. Published by Universitas Muhammadiyah Malang
This is an open access article under the CC BY SA license. (https://creativecommons.org/licenses/by-sa/4.0/)
Kelantan river basin covers an area of about 13,000 km² together with its tributaries; they are Lebir river, Galas river, Pergau river, and Nenggiri river [21]. The Kelantan river is approximately 105 km and it includes Lebir river and Galas river in Kuala Krai city. The Lebir river and Galas river are the central part of the Kelantan river. Respectively its about 2,430 km² and 7,770 km² [22]. Figure 1 shows the river network of the Kelantan watershed, major cities, and water level stations. The total length of the main river, Kelantan, is about 388 km from the head of its longest tributary and drains an area about 13,000 km² occupying more than 85% of the Kelantan State [16].

The river water level data is acquired from the Department of Irrigation (DID) Malaysia with fifteen minutes basis. About twenty-seven months of data in January 2013 – March 2015 were collected through DID supervisory control and data acquisition systems, used as dataset in this study. It is about 78688 records datasets were used, employed for training and validation test. Three variables as shown in Table 1. and Figure 1 including four multi-input variables of each feature indicates the river water level, rainfall, and river stream flows. These input data with limited physical experiences are required for ANFIS network, and one observed water level as the target of output.

2.2 Adaptive Neuro-Fuzzy Inference System (ANFIS)

If the input data are ambiguous or subject to relatively high uncertainty, a fuzzy system such as ANFIS may be a better option. Jang [23] presented the Adaptive Neuro-Fuzzy Inference System (ANFIS) which was obtained from ANN and Fuzzy logic by catching the advantages of both in one framework. The ANFIS is a multilayer feedforward network which uses neural learning algorithms and fuzzy reasoning to map an input space to an output space. With this ability of a neural system adaptive network, ANFIS has been shown to be powerful successfully used in diverse fields at solving non-linear issues and optimization problems [24].

This system is a fuzzy Sugeno by a forwarding network structure and will be employed in this study. This models are typically developed and placed into the framework of a neural network model into enable adaptation [25]. Considering an example in this process of a fuzzy inference system with two inputs, five layers of ANFIS and one output as can be seen in Figure 2. The ANFIS network consist of six layers such as input layer, fuzzification layer, inferences process in layer-3 and layer-4, defuzzification layer, and inference layer as final output. Input factor from layer-1 is fuzzified by layer-2. And then inferences process and rules including strength normalization are done in layer-3 and layer-4, respectively. Calculation of output for each corresponding rules are carried out in layer-5 which is known as defuzzification. In final output layer, this layer is single node which calculates the total number of all received signals from layer-5. The functions of each layers of ANFIS network as generally presented in [25].

2.3 Performance Evaluations

In order to demonstrate the efficiency of the proposed models, both of the individual and CMIS conceptual models will be evaluated in term of a series of different performance measures. Four different quantitative performance indicators will be evaluated; root mean square error (RMSE), mean absolute error (MAE), mean percentage error (MPE) and correlation coefficient (R). Detailed formulations can be found in [5] and [7]. RMSE indicates overall agreement in

---

**Table 1. Input and Output Configurations Used for ANFIS Flood Forecasting Model**

<table>
<thead>
<tr>
<th>Water Level</th>
<th>Rainfall</th>
<th>Streamflow</th>
<th>River Water Level Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>WL1-Sg. Galas at Chegar Lapan</td>
<td>RF1-Bertam</td>
<td>SF1-Sg. Galas at Chegar Lapan</td>
<td>Sungai</td>
</tr>
<tr>
<td>WL2-Sg. Sokor at Tegawan</td>
<td>RF2-Kuala Nal</td>
<td>SF2-Sg. Sokor at Tegawan</td>
<td>Kelantan at</td>
</tr>
<tr>
<td>WL3-Sg. Lebir at Tualan</td>
<td>RF3-Lebir</td>
<td>SF3-Sg. Lebir at Tualan</td>
<td>Guillemand</td>
</tr>
<tr>
<td>WL4-Sg. Lanas at Air Lanas</td>
<td>RF4-Gemang Bahru</td>
<td>SF4-Sg. Lanas at Air Lanas</td>
<td>Bridge</td>
</tr>
</tbody>
</table>

---

**Figure 2. The Architecture of ANFIS for Two Input Problems According to [23][25]**

---

actual units with the deviations are squared so the assessment is biased in favor of higher magnitude events. MAE is indicates overall agreement in actual units. The relative error is a relative measure of the error relative to the observed record while \( R \) value describes the proportion of the variance in the observed dataset that can be explained by the model.

3. Data Pre-Processing and Model Development

The overall methodology used in the present study is shown in Figure 3. In which raw data were retrieved from river water level, rainfall, and streamflow in the upstream area, while one label output from downstream river water level as FFP. These input and output configurations involved each different stations as described in Table 1. The dataset were divided into two parts namely, training set and test (validation) set.

An analytical definition of a flood water level forecasting model \( Y \), considering \( X_n \) multivariable inputs data including river water level \( Y \), streamflow \( Q \), and rainfall \( R \), can be expressed as Equation 1. Machine learning based method utilizing ANFIS model is used to estimate the hydrological transfer function \( \varphi(.) \) to characterize the complicated of non-linear mapping relationship in a river basin including water levels and flows, and the rainfall factor [26].

\[
Y(t) = \varphi(Q(t - d_1 + 1), R(t - d_2 + 1), W(t - d_3 + 1))
\]  

(1)

The observed event-based water level data, rainfall, and streamflow were split into training and testing sets where fifteen months of the available data (twenty-seven months) is used for training while the remaining twelve months were used as test data. Data normalization is necessary for doing data pre-processing and better prediction. Data pre-processing is done to minimizing noises of data, detecting trends, and flatten the distribution of the variables. In this study, the normalization inputs and output data linearly between 0 to 1 were applied. A principal component analysis is involved to quantifying the importance of the hydrological parameters that effect the appearance of extreme events [27].

Table 1 shows the number of the configuration of input vectors uses in the ANFIS model simulation, which are involves of three features including water level, rainfall, and streamflow. In which, each features consists of four multiple input variables, respectively for water level-1 to water level-4, rainfall-1 to rainfall-4, and streamflow-1 to streamflow-4. One output vector as a target is considered as FFP to be predicted. These input and output variables are represented their station identification number and location name [28].

The combination of least-squares and backpropagation gradient descent methods as training algorithm is used to model the training data set. The related function in MATLAB© namely \textit{genfis} and its parameters were selected to simulate the ANFIS model. The related function in MATLAB© namely \textit{anfis}, then used to generate a single output FIS structure from the training data set and tunes the system parameters using specified input/output training data. The number and type of membership functions were set by defaults. While the option number of epochs was set to 10.

4. Results and Discussion

Figure 4 illustrates the actual hourly river water level value, respectively for 1-hr ahead, 5-hr ahead, and 12-hr ahead of lead time forecasting. These figures indicate that the model could follow the peak value of river water level as well as the flood water level could be forecasted employing the proposed ANFIS model. While Figure 5. Shows the forecasting accuracy and efficiency of the employed ANFIS model in term of various evaluations criteria. The short-term period lead forecasting could result the lowest error. In addition, the lowest error indices the best accuracy of the model performance. Likewise, the higher of forecasting time ahead period, the performance of the model would be getting worst [29].
As shown in Figure 5(a) and Figure 5(b), the error performance values of the forecasting period in both training and testing show an increasing trend with the increases of lead-time. However, among the error performance evaluation, result on MAE has shown in the lowest value. Therefore, Table 2 shows result on R-performance shows a decreasing trend followed by increases of lead-time of forecasting. It can be seen in Figure 5(c). Furthermore, results on both MPE and MAE in testing data were outperformed with minimum value over the training phase. In addition, in term of R-performance, the result on testing periods (test set) produced high value and leads outperform over the training periods. However, only RMSE results were produced higher value in most multi-time ahead forecasting (except 7 and 8 hour ahead) in the testing phase against the result value in the training phase. In this case, Choubin [30] describes that ANFIS model could overlearn during the training period, which it could leads to reducing the performance during the testing period.

<table>
<thead>
<tr>
<th>Lead Time</th>
<th>MPE</th>
<th>RMSE</th>
<th>MAE</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Test</td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>1 hr</td>
<td>0.1282</td>
<td>0.1147</td>
<td>0.0221</td>
<td>0.0358</td>
</tr>
<tr>
<td>2 hr</td>
<td>0.2411</td>
<td>0.2008</td>
<td>0.0392</td>
<td>0.0430</td>
</tr>
<tr>
<td>3 hr</td>
<td>0.3622</td>
<td>0.2973</td>
<td>0.0567</td>
<td>0.0609</td>
</tr>
<tr>
<td>4 hr</td>
<td>0.4905</td>
<td>0.4016</td>
<td>0.0748</td>
<td>0.0830</td>
</tr>
<tr>
<td>5 hr</td>
<td>0.6251</td>
<td>0.5110</td>
<td>0.0939</td>
<td>0.1075</td>
</tr>
<tr>
<td>6 hr</td>
<td>0.7632</td>
<td>0.6141</td>
<td>0.1136</td>
<td>0.1169</td>
</tr>
<tr>
<td>7 hr</td>
<td>0.9034</td>
<td>0.7100</td>
<td>0.1338</td>
<td>0.1175</td>
</tr>
<tr>
<td>8 hr</td>
<td>1.0443</td>
<td>0.8085</td>
<td>0.1544</td>
<td>0.1272</td>
</tr>
<tr>
<td>9 hr</td>
<td>1.1814</td>
<td>0.9492</td>
<td>0.1752</td>
<td>0.1718</td>
</tr>
<tr>
<td>10 hr</td>
<td>1.3152</td>
<td>1.1050</td>
<td>0.1963</td>
<td>0.2578</td>
</tr>
<tr>
<td>11 hr</td>
<td>1.4437</td>
<td>1.2495</td>
<td>0.2173</td>
<td>0.3393</td>
</tr>
<tr>
<td>12 hr</td>
<td>1.5668</td>
<td>1.3667</td>
<td>0.2382</td>
<td>0.3800</td>
</tr>
</tbody>
</table>

Figure 4. Sample Results of Hourly Representative Water Level Forecasting. a) 1-hr Ahead Forecasting, b) 5-hr Ahead Forecasting, and c) 12-hr Ahead Forecasting

The regression coefficient $R$ was also calculated and plotted as shown in Figure 6, the correlation coefficients are adequate with more than 90% even though for 12-hr-ahead forecasting. If the forecasted and actual river water being compared are similar, the scatter points should approximately lie on the regression line $y = x$ [26]. It indicates there was a significant positive correlation between the forecasted and actual value. The high value of $R$ presented in Figure 6 also described the accuracy of the ANFIS model employed in this hydrological problem especially for river water level forecasting for flood warning.

Small sample sizes have been a serious limitation for many earlier studies. It is agreement as suggested by the work presented in [19], the more input variable could reflect the performance of the model. It is proved that there is an improvement of the result produced by ANFIS in term of $R$ more input variables including rainfall and streamflow investigated in the present study. It shows in Table 3 with more input configurations and limited physical knowledge of attributes presented an adequate result performance of the model against previously published works.

![Figure 5. Error Performance Result of the Training Period and Testing Period](image1)

![Figure 6. Sample of Hourly Coefficient of Correlation Performance of Actual Flood Water Level against ANFIS Forecasted Model. (a) 1hr-Ahead, (b) 5hr-Ahead, and (c) 12hr-Ahead of Time Forecasting](image2)

<table>
<thead>
<tr>
<th>Input variables</th>
<th>Model</th>
<th>RMSE</th>
<th>MPE</th>
<th>MAE</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>river water level</td>
<td>RBFNN</td>
<td>0.1439</td>
<td>-</td>
<td>-</td>
<td>0.9752</td>
</tr>
<tr>
<td>[19]</td>
<td>NARX</td>
<td>0.1154</td>
<td>-</td>
<td>-</td>
<td>0.9836</td>
</tr>
<tr>
<td>rainfall, streamflow</td>
<td>SVM</td>
<td>0.2843</td>
<td>-</td>
<td>-</td>
<td>0.9864</td>
</tr>
<tr>
<td></td>
<td>ANFIS</td>
<td>0.3800</td>
<td>1.3667</td>
<td>0.1487</td>
<td><strong>0.9902</strong></td>
</tr>
</tbody>
</table>


4.1 Flood Disaster and Early Warning

In connection with disaster risk management and flood warning system, according to DID [31] and [32] for river water level data-above sea level, it can be classified as three main categories, including “alert level”, “warning level”, and “danger level”. The alert level for Kelantan River, in Kelantan is 12.50 meters, while for warning level is 15.10 meters, and danger level is 18.60 meters – above sea levels [33].

Figure 7. An Illustrated Time Sequences of Flood Forecasting, Flood Warning, and Responses Lead Times

Figure 7 illustrated a time sequences of flood forecasting, flood warning, and responses lead times. Occurrences of heavy rainfall would flows into the reservoir and are expected to increase in river water level in just a couple of hours. It indicates when the river water level more than the maximum observed value, can be assumed flood will occur in that particular area since it could reach at a danger level.

From Table 4, it can be observed that ANFIS model presented an adequate result in order to build flood early warning and disaster risk reduction. The average value between actual and ANFIS model showed precisely which these results are in accordance with normal condition in green level. On the other hand, ANFIS model also performed well in highest value forecasting, which this value is exceeded the three risk categories as mentioned. This condition may lead flood disasters occurred. As an early warning system this flood forecasting by ANFIS model is effective tools to detect flood water level at the certain levels.

Table 4. Kelantan River Water Level Threshold for Flood Warning

<table>
<thead>
<tr>
<th>Normal Level (Green)</th>
<th>Alert Level (Yellow)</th>
<th>Warning Level (Orange)</th>
<th>Danger Level (Red)</th>
<th>Average Level</th>
<th>Highest Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td>ANFIS</td>
<td>Actual</td>
<td>ANFIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.10 M</td>
<td>12.50 M</td>
<td>15.10 M</td>
<td>18.60 M</td>
<td>9.09 M</td>
<td>9.05 M</td>
</tr>
<tr>
<td>22.78 M</td>
<td>21.34 M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Conclusion

Flood forecasting technique reflecting historical data of the upstream river basin to predict the flood water level in the downstream area has demonstrated in this study. The Kelantan river, located in Kelantan State, Peninsular Malaysia was taken as a case study. These upstream and downstream river basins respectively used as features input and output configurations, consists of multiple input variables for each feature including river water level and flows, and also rainfall intensities. They were considered in the model structure design. ANFIS model is established to forecast river water level for different lead times in 1-hr to 12-hr ahead. Their performances were evaluated by error performance analysis and correlation coefficient result.

The ANFIS model demonstrated good performance for flood water level forecasting, and the R performance decreases with the increase of lead-times. While the error values performed increases with the increase of lead-times. The ANFIS model performed well compared with RBF, NARX and even with SVM with produced 0.9902 in 12-hr ahead of time forecasting. Highlighted that these results were produced by comprising more input variables to improve the performance of the model. Although some machine learning models has attempted, ANFIS model produced more promising results and can be used as a practical tool for flood forecasting technique with limited physical knowledge in both short-term and long-term basis.

These findings make several contributions to the current literature intelligent frameworks in order to build a committee machine with intelligent systems (CMIS) developing by the present authors. These individual learning machines could improve the performance of the model to get the generalization and robustness of the flood forecasting technique. In the context of advanced computational methods, for future research work, CMIS could also help as a promising optimization tool in the hydrological time-series forecasting topics.

Notation

\[ Y(t) \] : the forecasted river water level at time \( t \).

\[ Q(t - d_1 + 1) \] : the previous streamflow up to \( t - d_1 + 1 \) time steps.

\[ R(t - d_2 + 1) \] : the rainfall data with \( t - d_2 + 1 \) time steps.

\[ W(t - d_3 + 1) \] : the water level data with \( t - d_3 + 1 \) time steps.

\( \psi(\cdot) \) : hydrological system transfer function performed by ANFIS model.

\( d_{i1} \) : the length of time ahead (hour).

Acknowledgement

This work is performed using the Japan-ASEAN Integration Fund (JAIF) with reference number UTM.K43/11.21/1/12 (264) year 2018-2021, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia. Dataset used in this research were provided by Department of Irrigation and Drainage (DID), Malaysia. The first author would like to acknowledges Universitas Muhammadiyah Malang for giving the opportunity to undertake this study.

References


© 2021 The Authors. Published by Universitas Muhammadiyah Malang

This is an open access article under the CC BY-SA license. (https://creativecommons.org/licenses/by-sa/4.0/)


