Flood vulnerability assessment using artificial neural networks in Muar Region, Johor Malaysia

A M Wahab and A N Muhamad Ludin

1 Student, Department of Urban and Town Planning, Faculty of Built Environment, Universiti Teknologi Malaysia
2 Professor, Centre for Innovative Planning & Development, Faculty of Built Environment, Universiti Teknologi Malaysia

E-mail: ahmadmubinwahab@gmail.com

Abstract. The measurement of Flood Vulnerability Assessment (FVA) is essential because it is the key towards disaster risk reduction as well as improving the community. Several existing techniques can be used to measure FVA, however, these processes are challenging, such as in determining the weighting factors. Therefore, this study attempts to apply Artificial Neural Networks (ANN) technique in the assessment of FVA because it can determine the weight between both input and output data through the network training process. The ANN technique has been widely applied in flood modelling; however, the application of ANN in FVA has not been explored extensively, mainly due to difficulties in obtaining real data. The purpose of this study is to assess the capability of Multilayer Perceptron (MLP) ANN technique with Lavenberg-Marquardt back-propagation algorithm in determining FVA in Muar region, Johor, Malaysia. The ANN architecture for this study is 9-N-N-1, i.e., nine nodes in the input layer, two hidden layers with N nodes, and one node for the output layer. In the training process of the network, 30 sampling locations with a 9x9 window were selected in the floodplain and non-flooded areas. Meanwhile, another 30 sampling locations were used to test the network in order to measure the performance model. Then, the trained network was used to generate FVA map for the study area. The result of the performance model shows the root mean square error is 0.0035 has been made between input data with target data. The findings of this study showed that the ANN technique can precisely approximate the FVA without predetermined weighting factors, but it depends on the target data used. The result of this study can also be used as a tool to assist decision makers in preparing national and local plans which take into consideration flood risk management.

1. Introduction
Flood disaster is one of the major catastrophic disasters in Malaysia. It often occurs when there is heavy rain that lasts for a few hours; causing the water to rise in level and overflow towards the residential area near the river. According to the Department of Irrigation and Drainage (DID), approximately 9% of the land area is vulnerable to flood, and 22% of the total population of the country live in flooded areas [1]. This disaster has not only cost many lives and destroyed properties, but also led to damage in the agricultural area and infrastructure, causing the country to incur a loss of approximately USD 0.3 billion annually [2]. In Johor, the worst flood occurred in December 2006 and...
January 2007, paralyzing several towns in the state; which resulted in a total of 110,000 victims being evacuated to flood relief centres, 18 lives claimed and estimated the loss of USD 0.5 billion [3]. In this case, the DID agency is responsible for implementing flood mitigation strategies in the country to ensure that the subsequent impacts of the floods can be mitigated through structural methods, primarily by using engineering methods [4]. However, the overall efficiency of flood management through this structural method is quite limited [5]. Therefore, the need for vulnerability assessment in flood management is required to provide accurate information so that the flood risk management in flood-prone areas can be implemented more efficiently.

Vulnerability assessment is a component of disaster risk assessment; the term was first introduced in the early 1980s by social scientists who questioned hazard-centric disasters [6]. In general, the vulnerability can be explained as the level of the system is susceptible and incapable of addressing a disaster by causing severe impact on environmental change [7]. Many studies have been conducted to determine whether the FVA has been carried out, mainly using general Flood Vulnerability (FVI) equation [8,9], Principal Component Analysis (PCA) [10,11] and Multi-Criteria Decision Analysis (MCDA) [12,13]. One of the essential processes in FVA is determining weight factors for each indicator and component; where it plays a vital role in determining different contributions for each indicator that affects FVA. It can be determined through several ways such as equal weighting [14], weighted linear combination [15] and expert judgment [12]. However, there are also some studies not using weighting in FVA calculations [8,9] because of the number of different judgment ratings lying behind combined or interpolating weight [8]. The expert judgment technique can determine the weight of each indicator and component based on the actual situation, but it depends on the expert who gives the judgment rating.

ANN is a computational modelling tool that can solve complex correlation problems. In recent years, many hydrological studies have used ANN techniques in flood modelling because it can solve the problem of uncertainty in inputs and produce outputs from incomplete datasets [16]. Most of the studies conducted use rainfall and runoff parameters as the input and output without taking into account other factors that cause flood [3]. Therefore, the Kia et al. study [3] has used several flood causative factors as the input and the result of this study has shown an agreement between predicted and the real hydrological records. From this study, it shows that the ANN has the capability in flood modelling using several flood causative factors and the potential to be applied in FVA estimation.

This study aimed to explore the potential of ANN in estimating FVA in Muar region, Johor, Malaysia. The ANN technique used in this study is Multilayer Perceptron (MLP) with Levenberg-Marquardt (LM) backpropagation as a learning algorithm for training and testing the network.

2. Study Area
This study focused on Muar and Tangkak districts, where Muar river is the main river for both districts. The Muar River is located in Muar River Basin which flows through Negeri Sembilan, Pahang, and Johor, with a total distance of 329 km. The upstream part of this river covers the area of Kuala Pilah, Negeri Sembilan. Meanwhile, the downstream part covers the Muar and Tangkak districts; where the river mouth is located in Muar town and flows out to Malacca Straits. Muar River has experienced several significant floods occurring in 2006, 2007, 2011 and 2015, affecting nearly 10,000 residents living in the Muar catchment area [17].

Agriculture dominates the entire study area (69.5%), followed by natural forest (20.8%), urban areas (6.7%), water body (2.6%) and grassland (0.4%). The main crops for this area are palm oil (54.3%), rubber (29.8%), fruits (9.3%) and other crops (6.6%). The highest density area in Muar district is Muar town with a total population of 94,929 people; while in Tangkak town, the total population is 51,555 people [18]. Figure 1 shows the map for the study area, where the blue striped area indicates the flood-prone areas. From this Figure, it shows that a total of 60 sampling points are randomly divided into two categories: training sampling points (orange box) and validation sampling points (yellow box).
3. Methodology
Overall, the method used for this study is through three major processes that involve the formulating of a conceptual framework, selection of indicators, and FVA estimation using ANN as shown in Figure 2. The conceptual framework for FVA estimation in this study is based on the conceptual framework developed by a team of experts from Adelphi and European Academy of Bozen (EURAC) in 2004 [19], which can be seen in Figure 3. From this figure, the conceptual framework is divided into four components of exposure (E), sensitivity (S), potential impact (PI) and resilience (R) which includes physical, environmental and social dimensions.
3.1. Vulnerability indicators

For the selection of indicators, nine indicators are chosen and classified into components of exposure, sensitivity, and resilience based on their characteristics by the conceptual framework developed. Table 1 shows the list of the selected vulnerability components with their factors, indicators and data sources that are used to estimate FVA using ANN. These indicators are derived on a spatial size of 10 x 10 meter into the raster layer, respectively. Then the values for these indicators are normalized to a range of 0 to 1 using the following formula:

$$z_i = \frac{(x_i - x_{min})}{(x_{max} - x_{min})}$$  \hspace{1cm} (1)

Where $z_i$ represents normalized value, $x_i$ as the individual value, $x_{min}$ as the lowest value, and $x_{max}$ as the highest value.

| Vulnerability Component | Factor | Indicator | Data |
|-------------------------|--------|-----------|------|
| Exposure                | Precipitation | Average daily rainfall over five years (mm/day) | Rainfall station with TRMM\(^a\) distribution |
|                         | Topography | Elevation (m) | Alos-Palsar\(^b\) DEM\(^c\) |
|                         |          | Slope gradient (in degrees) | Alos-Palsar\(^b\) DEM\(^c\) |
| Sensitivity             | Land use | Classified land cover LULC\(^d\) map | map |
|                         | Soil type | Classified soil type map | Soil type map |
|                         | Main river | Buffered main river area | River line features |
|                         | Population | Population density per grid cell | National census 2010 |
| Resilience              | Road | Access to road | Road network |
|                         | Drainage | Buffered drainage area | Drainage line features |

\(^a\) Tropical Rainfall Measuring Mission.  
\(^b\) Advanced Land Observation Satellite - Phased Array type L-band Synthetic Aperture Radar.  
\(^c\) Digital Elevation Model.  
\(^d\) Land use/Land cover
3.2. Flood vulnerability assessment using ANN

This study uses the ANN multi-layer perceptron (ANN-MLP) structure that consists of the three-interconnection layer of the input layer, hidden layer, and an output layer. The ANN-MLP is often used in hydrology because it can estimate any function with a finite amount of deficiency [20]. Figure 4 shows the ANN structure used, which consists of nine neurons in the input layer, 18 neurons in the first hidden layer, nine neurons in the second hidden layer, and one neuron in the output layer. Two hidden layers were selected because the use of more than one hidden layer showed greater flexibility than using a single hidden layer [21]. Furthermore, several past studies have used two hidden layers as the starting point for their researches [22].

![Figure 4. ANN architecture](image)

In the data preparation process, the sample of 30 boxes of the 9x9 window (see Figure 1) were randomly selected to use in the process of training the network; while another 30 sample areas were selected for the validation process. For the selected area samples, they are distributed equally and located in flood prone areas and non-flood prone areas. The target data or desired output is produced using multi-criteria evaluation method with fuzzy logic based on research done [15] in the area of Iskandar Malaysia, Johor. Then, the input and target data are compiled into a data layer: the input is arranged into each row representing each indicator, and the target data is arranged into a single row.

The next process is to determine the ANN configuration by setting the value of parameters as shown in Table 2. Then, the network is trained to determine the weight and bias for each network by changing the value of the weight between the input and output layer so that the minimum error can be achieved by using the Lavenberg-Marquardt back-propagation algorithm. After the network is trained, the validation of the network is performed to test the efficiency of the network using input and target data in the new sample area. Finally, the FVA for the entire study area is produced by using the network that has been trained and validated.

| Parameters                      | Value                                |
|---------------------------------|--------------------------------------|
| Number of input data            | 9                                    |
| Number of hidden layers         | 2                                    |
| Number of output data           | 1                                    |
| Training function               | Levenberg-Marquardt back-propagation |
| Training parameter minimum gradient | 0.001                              |
| Training parameter epochs       | 1000                                 |
| Training parameter learning rate | 0.4                                  |
| Training parameter max. fail    | 20                                   |

Table 2. List of ANN configurations.
4. Result and Discussion

In the present study, the data assessment is conducted in two ways—network performance assessment and map assessment. The process of network performance assessment is carried out to test the accuracy between ANN’s output and target data. The result of this process is presented through three different plots which are training state plot, validation performance plot, and linear regression plot, as shown in Figure 5-7. For the map evaluation process, Figure 8-9 shows the spatial variability of FVA to assess the different levels of vulnerability based on the different intensities and classes of FVA.

4.1. Network performance assessment

Figure 5-6 shows the graph for the training state plot and the performance plot during the training phase using the Levenberg-Marquardt back-propagation algorithm technique. From Figure 5, the three different plots represent three parameters for validation check, i.e. gradient versus epochs, Momentum Update (MU) versus epochs, and validation fails versus epochs. It shows that the algorithm has truly converged at the minimum gradient parameter of $9.9131 \times 10^{-8}$, mu parameter of $1 \times 10^{-8}$, and there is no validation fail at 70th epoch with Mean Squared Error (MSE) of $1.8596 \times 10^{-9}$. Figure 6 indicates that a small value of MSE nearing zero determines that the ANN’s output and the target data are well trained.

![Figure 5. Training state plot](image)

![Figure 6. Validation performance plot](image)

![Figure 7. Linear regression relation between ANN’s output and target data](image)
In order to check the performance assessment, the ANN’s output is compared with the target data using linear regression plot, which can be seen in Figure 7. The parameters evaluated in this performance assessment are determination coefficient ($R^2$) and root mean squared error (RMSE). The $R^2$ value obtained is large ($R^2 = 0.996$), which indicates that this model is suitable for estimating FVA. Meanwhile, the error value obtained for this comparison is 0.0035, showing a minimal error in the RMSE value. This comparison result shows that the FVA estimated by the ANN-MLP technique has higher accuracy when compared with the target data. It proves that the ANN-MLP technique has the potential to estimate FVA based on input and target data provided.

4.2. FVI map using ANN.

Figure 8-9 shows FVA intensity map and FVA class map, which are produced using the ANN-MLP technique in the areas of Muar and Tangkak, Johor. The FVA intensity map (Figure 8) shows the spatial variability of FVA with a range between 0 to 0.85 (low to high). In Figure 9, the FVA is classified into five classes (very low, low, medium, high and very high) based on the equal interval range.

![Figure 8. FVA intensity map](image_url)

![Figure 9. FVA class map](image_url)

From both maps, the more vulnerable areas ($> 0.8$) can be seen with the following characteristics: low-lying areas ($< 30$ m), residents living close to the river ($< 8.3$ km), and high population density ($> 10$ people per grid cell of 10 km). While the lower vulnerability ($< 0.4$) area is characterized as follows: high topography areas ($> 90$ m), residents living away to the river ($> 16$ km), and low population density ($< $ one people per grid cell of 10 km). According to Figure 9, the dotted black circle indicates that the degree of vulnerability for this area is very high ($> 0.8$). This area is a residential area in Muar Town that has a higher population density compared to other areas, as located in low-lying areas and close to the river.

The three significant factors that influence the difference in the level of FVA are topography, the distance of residents between the river and the population density. Other indicators such as rainfall, land use, soil types, roads and drainage also affect the vulnerability level of a particular area, but the effects are not as significant. The results of this study show that the ANN-MLP techniques can determine the vulnerability of the floods, where it is more efficient in determining the weighting factors as discussed earlier.
5. Conclusion
In this study, the ANN-MLP technique with Lavenberg-Marquardt back-propagation algorithm as the training function is used to determine FVA for Muar and Tangkak district, Johor. It is noted that the FVA is successfully estimated using this technique through several processes providing inputs and target data, configuring ANN parameters, training, and validation ANN, and producing FVA map using ANN's trained. The results from the network performance assessment indicate that small errors (0.0035) have been achieved between ANN's output and target data. Furthermore, the FVA map shows that high FVA areas are located in low elevation areas, close to the river and high population density areas. This FVA map can be used as a necessary data to assist flood risk management in reducing flood risk in this country.

Acknowledgements
The authors would like to thank Faculty of Built Environment and Universiti Teknologi Malaysia (UTM) for giving support to conduct this research. Special thanks also go to the Ministry of Higher Education (MoHE) for providing financial under the MyBrain15 scholarship. Lastly, we would like to show our gratitude to the Department of Irrigation and Drainage Malaysia, Department of Statistics Malaysia, and Department of Town and Country Planning Malaysia for providing data and information that greatly assisted the research.

References
[1] Department of Irrigation and Drainage Malaysia 2009 Flood Management Manual vol 1
[2] Pradhan B 2010 Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing J. Spat. Hydrol. 9
[3] Kia M B, Pirasteh S, Pradhan B, Mahmud A R, Sulaiman W N A and Moradi A 2011 An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia Environ. Earth Sci. 67 251–64
[4] Asmara T A T and Ludin A N M 2014 Mapping Perception of Community Preparedness towards Flood in Muar River, Johor Malaysia IOP Conference Series: Earth and Environmental Science vol 18(IOP Publishing)p 12126
[5] Liu P S and Chan N W 2003 The Malaysian flood hazard management program Int. J. Emerg. Manag. 1 205–14
[6] Prospero J M 1999 Long-term measurements of the transport of African mineral dust to the southeastern United States: Implications for regional air quality J. Geophys. Res. 104 15917
[7] Change C 2001 The Scientific Basis. Intergovernmental Panel on Climate Change by JT Houghton, Y. Ding, DJ Griggs, al
[8] Balica S F, Douben N and Wright N G 2009 Flood vulnerability indices at varying spatial scales Water Sci. Technol. 60 2571–80
[9] Musa Z N, Popescu I and Mynett A 2014 The Niger Delta’s vulnerability to river floods due to sea level rise Nat. Hazards Earth Syst. Sci. 14 3317
[10] Fekete A 2009 Validation of a social vulnerability index in context to river-floods in Germany Nat. Hazards Earth Syst. Sci. 9 393–403
[11] Kazmierczak A and Cavan G 2011 Surface water flooding risk to urban communities: Analysis of vulnerability, hazard and exposure Landsc. Urban Plan. 103 185–97
[12] Kienberger S 2012 Spatial modelling of social and economic vulnerability to floods at the district level in Búzi, Mozambique Nat. Hazards 64 2001–19
[13] Lee G, Jun K S and Chung E-S 2015 Group decision-making approach for flood vulnerability identification using the fuzzy VIKOR method Nat. Hazards Earth Syst. Sci. 15 863–74
[14] Zachos L G, Swann C T, Altinakar M S, McGrath M Z and Thomas D 2016 Flood vulnerability indices and emergency management planning in the Yazoo Basin, Mississippi Int. J. Disaster Risk Reduct. 18 89–99
[15] Yeganeh N and Sabri S 2014 Flood vulnerability assessment in Iskandar Malaysia using multi-criteria evaluation and fuzzy logic Res. J. Appl. Sci. Eng. Technol. 8 1794–806
[16] Anctil F and Tape D G 2004 An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition J. Environ. Eng. Sci. 3 S121–8
[17] Bin Wan Abdul Majid W, Brown E, Bin Osman S, Bin Asan G, Bin Osman A, Bin Samsudin R, Boelee L and Ahmad F 2017 Flood forecasting and warning for Muar River: non-structural measures for flood mitigation
[18] Malaysia D of S 2010 Population Distribution and Basic Demographic Characteristic Report 2010
[19] Fritz sche K, Schneiderbauer S, Bubeck P, Kienberger S, Buth M and Zebisch M 2014 Kahlenborn, Walter-The Vulnerability Sourcebook-Concept and guidelines for standardised vulnerability assessments Ger. adelphi, EURAC-Institute Appl. Remote Sensing, Dep. Geoinformatics–Z_GIS, Univ. Salzbg. 33
[20] Hornik K, Stinchcombe M and White H 1989 Multilayer feedforward networks are universal approximators Neural networks 2 359–66
[21] Sarle W S 1994 Neural networks and statistical models
[22] Tamura S and Tateishi M 1997 Capabilities of a four-layered feedforward neural network: four layers versus three IEEE Trans. Neural Networks 8 251–5