International Journal of Engineering & Technology, 7 (4.19) (2018) 217-222

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Predicting the Occurrence of Landside at Penang Island, Malaysia, through Artificial Neural Networks Model

M.G.Hassan[1*], S.N. Othman[1], C.A.Taib[1], A.A. Ahmarofi[2], M.D. Akanmu[1], RosliZakaria@Che Kob[3]

[1]*School of Technology Management and Logistics, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia.*
[2]*School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah,Malaysia*
[3]*Pejabat Meteorologi Kedah, AlorSetar, Kedah*
**Corresponding Author Email: [1]ghozali@uum.edu.my*

Abstract

The occurrence of recurring landsidesthat plagued Penang Island in 2017 hasattracted much attention across Malaysia since the incident puts people's life in danger and worsensthe economy of state and country. Based on the harmful effects, the requirement for predicting the occurrence of landslide is deemed crucial for disaster management agencies to promote awareness and prepare for necessary action. Therefore, this study aims to investigate the meteorological factors, and finally to develop a predictive model based on Artificial Neural Networks (ANN). Two meteorological factors, which are daily mean temperature and daily rainfall are considered in this study based on current data from Malaysian Meteorological Department. As a result, only two percent misclassification rate is recorded by the model compared to the actual data,while its prediction performance is better than regression method. Besides, the correlation between the rangeof daily rainfall amount withcorresponding daily mean temperature for the occurrence of landslide is anovel contribution of this research as little attention had been given by previous studies. The developed ANN model is helpful in predicting the occurrence of landslide inPenang Island and provides beneficial guidance for the disaster management agencies to save properties and people's lives.

Keywords: *Landslide, Artificial Neural Networks, regression, Meteorological Factor, Data Mining Process.*

1. Introduction

The frequent occurrence of landslide at Penang Island throughout 2017 was reported as headlines in most main stream media across the nation. The worst incident,which created a panic situation among citizens,occurred on October, 2017 where 11 workers were killed at the construction site due tothe landslide [1]. Aside from the hilly areas in most of the state [2], it is found that Penang Island had received heavy rainfall level than usual during 2017 [3].Combined with the rainfall, the daily temperature during 2017 is considered high compared to the previous years starting from June, 2017 [4].However, to the best of our knowledge, there is lack of recent studies that relate the daily rainfall amount and daily mean temperature factors to the occurrence of landslides as found in [5], [6] and [7]. Hence, the correlation between heavy rainfall and high temperature during the period is explored in this research, particularlyin light of the contributionto the occurance of landslide in Penang Island.

Based on literature review, previous studies adopted various methods for prediction purposesuch as,fuzzy logic, case-based reasoning (CBR), support vector machines (SVM), regression analysis, decision trees and artificial neural networks (ANN). Fuzzy logic limitation lies in the difficulty in supplying membership information by experts or people [8][9], CBR faces issues in representing case among different types of cases [9], while SVM is less compatible due to limitation in achieving true generalization [10] and development, training and testing of the method tend to be time consuming [8][9]. Moreover, regression analysis, which is a traditional forecasting method, shows low performance for data mining process [9][10], while decision trees are best applied when the number of classes is low [11] and for classification purpose [12]. Comparatively, in many cases, ANN shows capability in prediction due to its ability to capture the underlying structural relations among various variables during learning process [13][14]. Moreover, ANN is a nonlinear data driven and self-adaptive approach [9][13]. Therefore, ANN model for predicting the occurrence of landslide is developed in this study as the method has been proven to solve many prediction problems as presented by [13][14][15][16].

In the following section, research methodology is described in terms of case problem, data collection and development of ANN predictive model. Subsequently, the prediction results are discussed, while conclusion is elaborated in the final section.

2. Research Methodology

To fulfill the research objectives, the research activities are established, which consist of four levels namely, case problem, literature review, data collection and development of predictive model to predict the landslide occurrence. Furthermore, the related levels in the research activities are specified as phases to be carried out to achieve the whole objectives as shown in Figure 1. It further elaborates the structure of these research activities with selected methods to fulfil each particular research objective.Case problem and literature review on landslide were discussed in section 1, whereas data collection and the development of ANN model are elaborated in the next subsections.

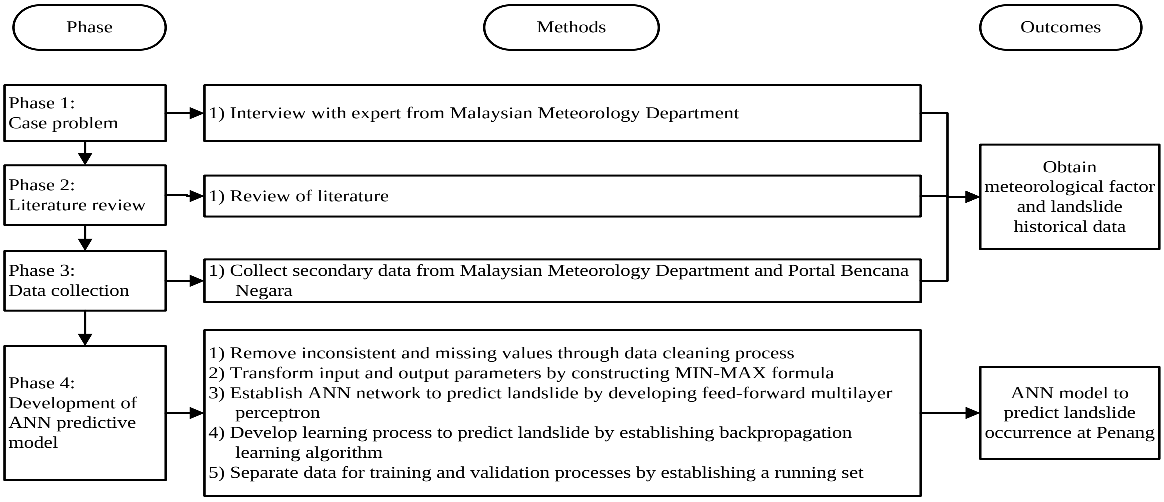

Fig. 1: Research Activities

2.1 Data Collection

In order to meet the objective of this research, data was collected to identify factors related to landslide occurrence. Data involved in this research are daily mean temperature and daily rainfall as demonstrated in Table 1. The related data was obtained from Bayan Baru, Penang Island records from 1st January, 2017 until 15th September, 2017. Data was collected from the reports of Malaysian Meteorology Malaysia. Table 1tabulates the source of secondary data from the organization.

Table 1: Source of secondary data

Secondary data	Source
Daily mean temperature in degree Celsius	Meteorological Malaysia
Daily rainfall in millimetre	Meteorological Malaysia

In addition, information and data on the occurrence of landslide in Penang Island during 2017 were collected from Portal Bencana Malaysia official website [17] (see Table 2).

Table 2: Date and location of landslide occurrences in Penang Island

Date of Occurrence	Location of landslide
14-September, 2017	Barat Daya, Penang Island
15-September, 2017	Barat Daya, Penang Island
20-September, 2017	BalikPulau, Penang Island
22-October, 2017	TanjungBungah, Penang Island

2.2 Development of ANN Predictive Model

In the development of ANN as a predictive model to meet the objective of this research, the collected data consisting of daily mean temperature and daily rainfall werecleaned through establishment of graph and normalization test to increase ANN prediction performance. Subsequently, related data were transformed as the input and output parameters through minimum and maximum (MIN-MAX) formula for ANN learning process as recommended by [14].
Subsequently, ANN feed-forward multilayer perceptron (MLP) network was developed for predicting the occurrence of landslide. After that, backpropagation (BP) learning algorithm was established for learning process as presented by [18] and [19]. Subsequently, the data was allocated between training and validation set during the learning process. Consequently, network with the smallest square error value waschosen to obtain the prediction of landslide occurrences. Figure 2 illustrates the flowchart for the development of ANN for predicting the occurrence of landslide.

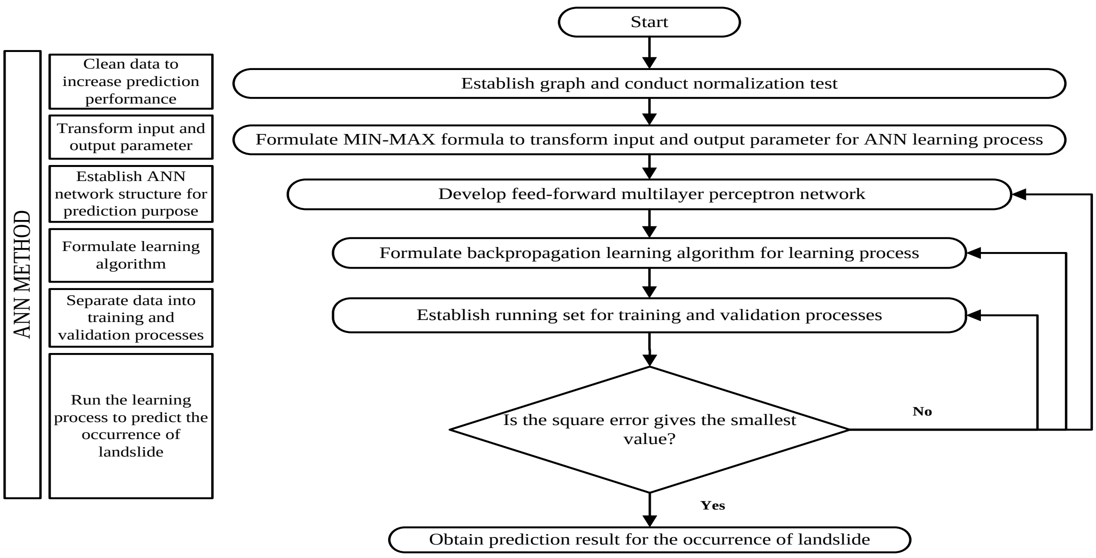

Fig. 2: Flowchart for development of ANN predictive model

2.2.1. Data Cleaning

Thedata of daily mean temperature and daily rainfall, which was obtained from Department of Meteorology Malaysia from 1st January, 2017 until 15th October, 2017 are plotted in a graph to observe any abnormality reading as presented in Figure 3 and Figure 4 as follows:

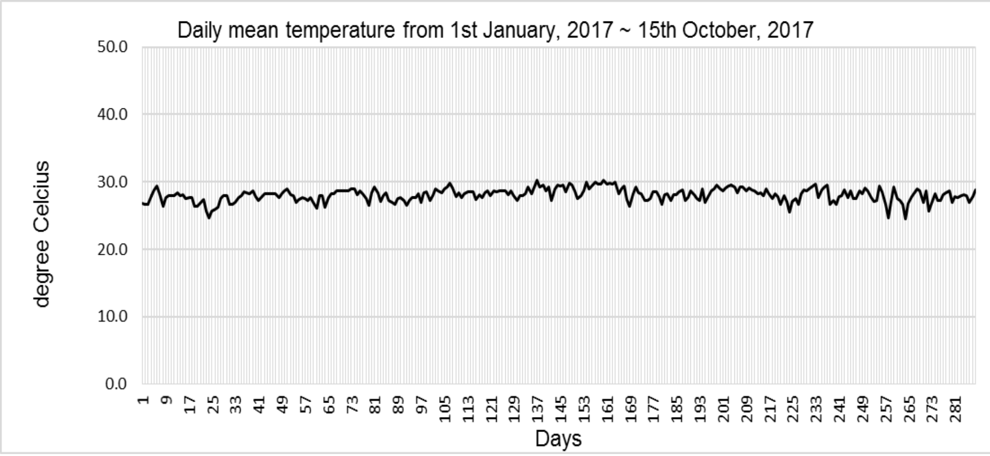

Fig. 3: Data for daily mean temperature

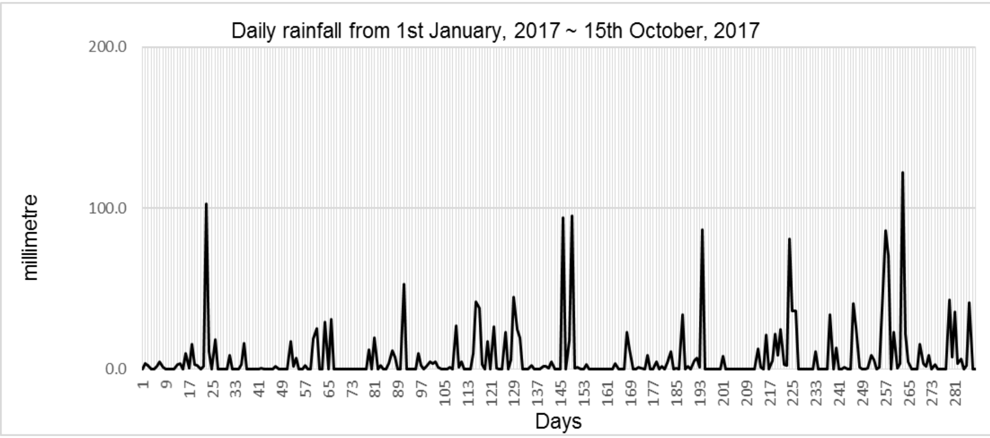

Fig. 4: Data for daily rainfall

From Figure 3 and Figure 4, the pattern of data for the daily mean temperature and daily rainfall shows normal fluctuation. Moreover, the related figures indicate a stationary pattern for increasing and decreasing plot. Thus, data has no abnormality and missing value, and is therefore acceptable to utilize in the ANN predictive model.

2.2.2. Transformation of Input and Output Parameter

The daily mean temperature, *temperature$_n$*, and daily rainfall, *rainfal$_n$*, are considered as input parameters, while the occurrence of landslide, *landslide$_n$*, is considered as output parameter. The input parameters (i.e., daily mean temperature and daily rainfall) are transformed between the interval of [0, 1] through MIN-MAX formulation as follows:

$$\mathrm{trans}_{\mathrm{temperature}_\mathrm{n}} = \frac{\mathrm{actemp}_\mathrm{n} - \mathrm{temp}_\mathrm{miv}}{\mathrm{temp}_\mathrm{mav} - \mathrm{temp}_\mathrm{miv}} \quad (1)$$

where,
$actemp_n$is the original degree Celsius of daily mean temperature for *n*th day
$trans_{temperature_n}$is the transformation value for degree Celsius of temperature for *n*th day
$temp_{miv}$is the minimum degree Celsius
$temp_{mav}$is the maximum degree Celsius

$$\mathrm{trans}_{\mathrm{rainfal}_\mathrm{n}} = \frac{\mathrm{actrain}_\mathrm{n} - \mathrm{rainfal}_\mathrm{miv}}{\mathrm{rainfal}_\mathrm{mav} - \mathrm{rainfal}_\mathrm{miv}} \quad (2)$$

where,
$actrain_n$is the original millimeter of daily rainfall for *n*th day
$trans_{rainfal_n}$is the transformation value for millimeter of rainfall for *n*th day
$rainfal_{miv}$is the minimum millimetre
$rainfal_{mav}$is the maximum millimetre
For output parameter (i.e., the occurrence of landslide,*landslide$_n$*), binary form is used to represent the output value. The occurrence of landslide is represented by 1 while no occurrence of landslide is represented by 0.

2.2.3. Development of ANN Network with Learning Algorithm

The learning process of BP learning algorithm is generated by the initialization of connection weight. Connection weight is initialized as the weight is a relative strength of neuron, which connects every single layer. The purpose of initializing connection weight value is to represent the importance of each input parameters (i.e. daily mean temperature and daily rainfall) towards desired output (i.e., occurrence of landslide) during the learning process. Therefore, MLP network is able to obtain the desired output value by correcting connection weight gradually and accordingly.

The value of connection weight is initially set to a random value, with no restriction on formulating the connection weight in a learning process as highlighted by [20]. The value of connection weight between *i*th input node and *j*th hidden node in the MLP network to predict landslide is formulated in the following equation:

$$\boldsymbol{w_{ab}} = \boldsymbol{W_{AB}} \quad (3)$$

where,
a is the number of input node at the input layer where $\forall \boldsymbol{a} = \boldsymbol{1}, \boldsymbol{2}, \dots, \boldsymbol{A}$
b is the number of hidden node at the hidden layer where $\forall b = 1,2, \dots, B$
Subsequently, the value of connection weight between *b*th hidden node and *c*th output node is formulated as follows:

$$\boldsymbol{w_{bc}} = \boldsymbol{W_{BC}} \quad (4)$$

where,
b is the number of hidden node at the hidden layer where $\forall b = 1,2, \dots, B$
c is the number of output node at the output layer where $\forall c = 1,2, \dots, C$
Furthermore, the formulation of summation function and sigmoid function of the input-hidden layer are expressed as follows:

$$sum_b = \sum_{a=1}^{A} w_{ab} trans_y \quad (5)$$

where,
sum_bisthe weighted sum of *b*th hidden node
w_{ab}isthe connection weight for the *a*th input node and *b*th hidden node
$trans_y$is the transformation value for respective input parameters where $\forall y = 1,2, \dots, Y$as
y=1for*temperature*$_n$ and *y*=2for*rainfal*
After that, the sigmoid transfer function of *b*th hidden node is formulated as follows:

$$sig_b = \frac{1}{(1 + e^{sum_b})} \quad (6)$$

where,
$sig_b$is the sigmoid value of *b*th hidden node
sum_bisthe weighted sum of *b*th hidden node
*e*is the base of natural logarithm with a constant value, 2.71828
Subsquently**,** the formulation of summation function and sigmoid function of the hidden-output layer are expressed as follows:

$$sum_c = \sum_{b=1}^{B} w_{bc} trans_z \quad (7)$$

where,
sum_cisthe weighted sum of *c*th output node
w_{bc}isthe connection weight for the *b*th hidden node and *c*th output node
$trans_z$is the transformation value for respective input parameters where $\forall z = 1,2, \dots, Z$as
z=1for*temperature*$_n$ and *z*=2for*rainfal*
The sigmoid transfer function of *c*th hidden node is formulated as follows:

$$sig_c = \frac{1}{(1 + e^{sum_c})} \quad (8)$$

where,
$sig_c$is the sigmoid value of *c*th hidden node
$sum_c$is the weighted sum of *c*th hidden node
*e*is the base of natural logarithm with a constant value, 2.71828
Figure 5 illustrates the flow of each step,comprising of connection weight, summation function and sigmoid function during the ANN learning process within the feed-forward MLP network to adjust connection weight.
The flow of the learning process in MLP network is generated in two directions, which are feed forward and back propagate. Both of the learning processes are iterated until the desired output of landslide occurrence is achieved. After that, summation function, sigmoid transfer function, and mean squared error are formulated to adjust the connection weight. Finally, the smallest value of square error, *E*, between the value of network output from sigmoid function and desired output is selected for predicting the occurrence of landslide. Thus, the connection weights is adjusted gradually and corrected accordingly to obtain the desired output.

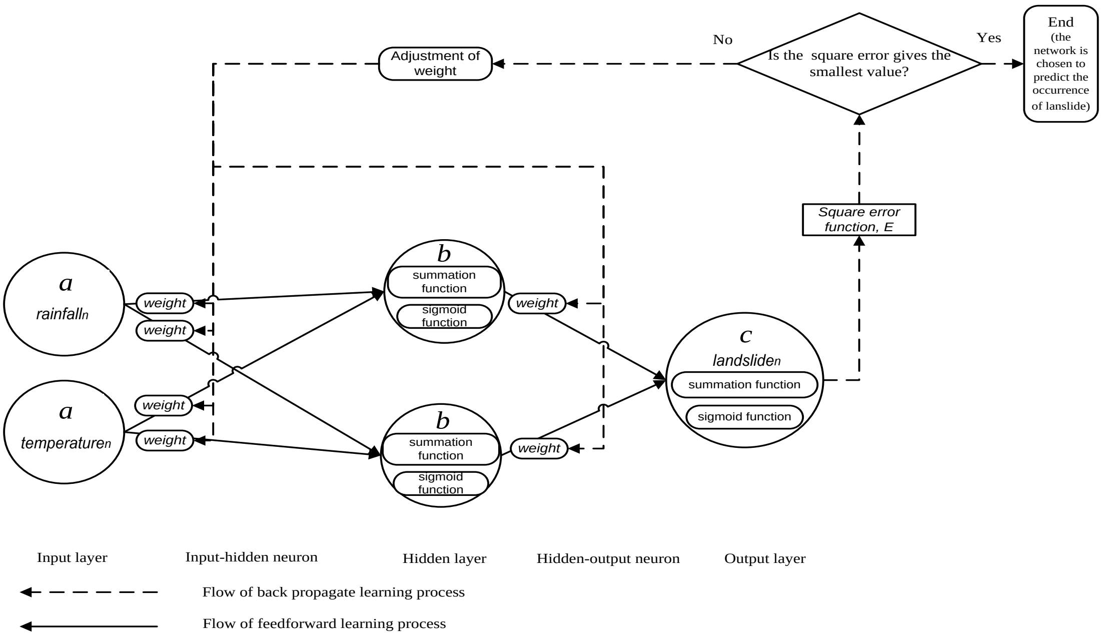

Fig. 5: The flow of learning process in the development of MLP network

Consequently, the predicted landslide occurrence is formulated in the following equation:

$$\mathrm{landslide_n} = [\mathrm{w_{ab}}]\mathrm{trans_{temperature_n}} + [\mathrm{w_{bc}}]\mathrm{trans_{rainfal_n}} + \sum_{\mathrm{b}=1}^{\mathrm{B}} [\mathrm{w_{bc}}]\mathrm{sig_b} \quad (9)$$

where,

$landslide_n$ is the predicted landslide occurrence of *n*th day

$[w_{ab}]$ is the final connection weight for *a*th input node and *b*th hidden node

$trans_{temperature_n}$ is the transformation value of *temperature* for *n*th day

$trans_{rainfal_n}$ is the transformation value of *rainfal* for *n*th day

$[w_{bc}]$ is the final connection weight for the *b*th hidden node and *c*th output node

sig_b isthe sigmoid value of *b*th hidden node

2.2.4. Data Allocation for Training and Validation Process

Training of ANN model is an iterative process during the learning process using a set of data until prediction error is minimum, while validation is a process to validate the model in predicting another set of data [9]. In this research, 80 percent of data was allocated to training process while 20 percent of data was allocated for validation process. By assigning more percentage of data for the training set, the ANN network gives better prediction performance since the more the data is trained, the stronger the predictive relationship it will have [14].

3. Results and Discussions

The input parameter (i.e., daily mean temperature and daily rainfall) for ANN predictive model is transformed between the interval of [0, 1] through MIN-MAX formulation as elaborated in subsection 2.2. Based on the Equation (1) and (2), the results of transformation value for daily mean temperature and daily rainfall are presented in the Table 3 for the first day, *n* =1.

Table 3: Transformation value of input parameters from formulated MIN-MAX

Input parameter	MIN-MAX formulation			Transformation value
	$actemp_1$	$temp_{miv}$	$temp_{mav}$	$trans_{temperature}$
$temperature$	26.8	24.5	30.2	0.4035
	$actrain_1$	$rainfal_{miv}$	$rainfal_{mav}$	$trans_{rainfal_1}$
$rainfal_1$	0.0	0.0	122.2	0.0

Subsequently, the result of prediction for the occurrence of landslide is indicated by observation number 8 and number 9 which is 1 (i.e., landslide has a possibility of occurring) as presented in Table 4. Both prediction results are aligned with reference to the two incidents that happened in September, 2017. Table 4 exhibits the prediction results for landslide occurrence based on daily mean temperature and rainfall.

Table 4: Prediction for occurrence of landslide based on daily mean temperature and rainfall

Observation	Daily mean temperature in degree Celsius	Daily rainfall in millimeter	Prediction results for the occurrence of landslide
1	27	4	0
2	26	19	0
3	26	25	0
4	27	12	0
5	30	1	0
6	29	0	0
7	28	0	0
8	27	86	1
9	25	71	1

Consequently, the prediction results of landslide based on range of daily rainfall amount and respected daily temperature are presented in the Table 5 as follows:

Table 5: Prediction for occurrence of landslide based on range of daily rainfall amount

Range of daily rainfall in millimeter	Daily mean temperature in degree Celsius	Prediction result for the occurrence of landslide
68 to 123	24	1
61 to 123	25	1
54 to 123	26	1
46 to 123	27	1
39 to 123	28	1
31 to 123	29	1
24 to 123	30	1

Based on the prediction result as presented in Table 5, the occurrence of landslide will likely occur when the temperature increases, while the rainfall amount decreases. The result is aligned with the fact that the higher the temperature, the more the soil moisture it will have [21]. Thus, the occurrence of the landslide is likely to happen as the soil moisture gets higher, caused by temperature and the related rainfall amount.

Furthermore, the misclassification rate for ANN model is 0.2 for validation process. Misclasification rate is aprediction error output of developed model compared to the historical data. Thus, the misclassification of two percent as recorded by the ANN model indicates that the prediction error for the model is considered low. Table 6 shows the misclassification rate obtained from the statistical report for the ANN predictive model.

Table 6: Misclassification rate for ANN predictive model

Statistic label	Validation process
Misclassification rate	0.2

Moreover, ANN predictive model has better prediction performance compared to regression method. The mean squared error of ANN predictive model is 0.005683, while it is 0.07227 for regression model.Table 7 presents the comparison of ANN predictive model and regression method in terms of mean squared error, *E*.

Table 7: Comparison of predictive method between ANN model and regression

Predictive method	Mean squared error
ANN	0.005683
Regression	0.07227

The predicitive method with smallest value of mean squared error indicates that the method is the best predictive model. Therefore, the ANN predictive model has better prediction and is recommended to predict the occurrence of landslide at Penang Island, based on the collected data.

4. Conclusion

A predictive model of ANN method based on MLP network and BP learning algorithm was developed in this paper to predict the occurrence of landslide at Penang Island. The learning process of ANN model to predict the occurance of landslide in the future was based on historical data, recorded throughout 2017. Two meteorological factors considered in this research are daily mean temperature and rainfall based on the landslidephenomenon that occurred in 2017. The results from the ANN predictive model indicated that the occurrence of landslide is likelywhen the daily temperature increases, while the rainfall amount decreases; for instance, when the range of daily rainfall for 24 Celsius is between 68 to 123 millimetre, and 24 to 123 millimetre for 30 Celsius. Moreover, the misclassification rate of ANN predictive model wasonly two percent, which is considered as low prediction error. Furthermore, prediction performance of ANN method is better than regression due to the fact that the mean squared error for ANN is less than that of the regression method. Therefore, ANN model is highly recommended to predict the occurrence of landslide at Penang Island.However, the topographical and geological factors are not considered in this research as the factors are beyond the scope of this research. Thus, both limitations serve as avenues for future work to improve the finding of landslide occurrence in Penang Island.

Acknowledgment

The Disaster Management Institute (DMI) UUM committee would like to thank the Research and Innovation Management Institute (RIMC) UUM for the research grant (code number 13734).

References

[1] "Kejadianberlakubegitupantas,"SinarHarian, p. 4, Oct. 22, 2017.
[2] B. Pradhann, H.J. Oh and M. Buchroithner,"Weights-of-evidence model applied to landslide susceptibility mapping in a tropical hilly area", Geomatics, Natural Hazards and Risk ,vol 1, no. 3, pp. 199-223,2010.
[3] Bernama,"Guan Engdedahkanpuncabanjir di Pulau Pinang",SinarHarian. Sept. 16, 2017. [Online] Available: http://www.sinarharian.com.my/edisi/utara. [Accessed Jan 29, 2018].
[4] Malaysia. JabatanMeteorologi Malaysia, Suhu Malaysia 2017. [Online] Available: http:// www.met.gov.my. [Accessed Jan 25, 2018].
[5] B.Pradhan and S.Lee, "Delineation of landslide hazard areas on Penang IslandIsland, Malaysia, by using frequency ratio, logistic regression, artificial nerual network models", Environmental Earth Sciences, vol. 60, no. 5, pp. 1037-1054, 2010.
[6] H. Lateh and J. Ahmad, J, "Landslide issues in Penang Island, Malaysia: Students' environmental knowledge, attitude and practice", Geografia-Malaysian Journal of Society and Space, vol. 7, no. 4, 2017.
[7] A.J Li, K. Lim, C.K. Chiu and B. Hsiung, "Application of Artificial Neural Network in Assessing Fill Slope Stability", World Academy of Science, Engineering and Technology, International Journal of Geotechnical and Geological Engineering, vol. 5, no. 2, 2018.
[8] N.C. Suresh, J. SlompandS. Kaparthi, "Sequence-dependent clustering of parts and machines: a fuzzy ART neural network approach",International Journal of Production Research, vol. 37, pp. 2793-2802, 1999.
[9] E. Turban, R. Sharda and D. Delen, Decision Support and Business Intellligence System.New Jersey: Pearson Education Inc., 2011.
[10] R. Carbonneau, R. Vahidov and K. Laframboise, "Machine learning-based demand forecasting in supply chains",International Journal of Intelligent Information Technologies, vol. 3, no. 4, pp. 40-57, 2007.
[11] K. Wang, "Applying data mining to manufacturing: the nature and implications",Springer Science and Business, vol.18, pp. 487-495, 2007, doi:10.1007/s10845-007-0053-5.
[12] C.F. Chien, W.C. Wang and J.C. Cheng, "Data mining for yield enhancement in semiconductor manufacturing and an empirical study",Expert Systems with Applications, vol.33, pp. 192-198, 2007.
[13] A.Azadeh, K.D. Shoushtari, M. Saberi and E. Teimoury, "An Integrated Artificial Neural Network and System Dynamics Approach in Support of the Viable System Model to Enhance Industrial Intelligence: The Case of a Large Broiler Industry", Systems Research and Behavioral Science, vol. 31, no. 2, pp. 236-257, 2014.
[14] A.A. Ahmarofi, R. Ramli, R and N.Z. Abidin, "Predicting Completion Time for Production Line in a Supply Chain System through Artificial Neural Networks", International Journal of Supply Chain Management, vol. 6, no. 3, pp. 82-90, 2017.
[15] A.K. Poonia, A.B. Soni and S. Khanam, "Optimization of operating parameters for sponge iron production process using neural network", International Journal of ChemTech Research, vol. 9,no. 2, pp. 20-34, 2016.
[16] N. Samattapapong and N. Afzulpurkar, "A production throughput forecasting system in an automated hard disk drive test operation using GRNN",Journal of Industrial Engineering and Management,vol. 9, no. 2, pp. 330-358, 2016.
[17] Malaysia. AgensiPengurusanBencana, ArkibBencana, BencanaTerdahulu, 2017. [Online] Available: http://portalbencana.ndcc.gov.my/Portal/Disaster. [Accessed Nov25, 2017].
[18] R.N.R. Dzakiyullah, B. Hussin, C. Saleh and A.M. Handani, "Comparison Neural Network and Support Vector Machine for Production Quantity Prediction", Advanced Science Letters, vol. 20 no. 10-11, pp. 2129-2133, 2014.
[19] R.N.R Dzakiyullah, "Production quantity estimation using improved artificial neural networks", M.S. thesis, UniversitiTeknikal Malaysia, Melaka, 2015.
[20] S. Samarasinghe, Neural Networks For Applied Sciences And Engineering: From Fundamentals To Complex Pattern Recognition. New York: Auerbach Publication, 2016.
[21] Y. Guo, J. Shi and K. Mao, K, " Surface temperature effect on soil moisture retrieval from AMSR-E", Geoscience and Remote Sensing Symposium, pp. 1192-1195, 2007.