The Implementation of The Neuro Fuzzy Method Using Information Gain for Improving Accuracy in Determination of Landslide Prone Areas

Winda Try Astuti[1], Much Aziz Muslim[2], Endang Sugiharti[3]

[1,2,3]Computer Science Departement, FMIPA, Universitas Negeri Semarang, Indonesia
Email: [1]windatryas@gmail.com, [2]a212muslim@yahoo.com, [3]endanghs02@yahoo.com

Abstract

The accuracy of information is increasing rapidly as technological development. For the example, the information in determination of disaster severity. The disasters that can be determined is landslide. This determination can be conducted using the fuzzy method. One of method is neuro fuzzy. Neuro fuzzy is a combined method of two systems, fuzzy logic and artificial neural network. The accuracy of neuro fuzzy method can be increased by applying the information gain. The purpose of this study is to implement and to know the accuracy of the implementation of information gain as the selection of landslide data features. It conducted to the neuro fuzzy method in determining landslide prone areas. The distribution of training data and testing data was using 20 k-fold cross validation. The implementation of the neuro fuzzy method on landslide data was obtained an accuracy of 81.9231%. In the implementation of the neuro fuzzy method with information gain was conducted in classification process. The process will stop when the accuracy has decreased. The highest accuracy result was obtained of 88.489% by removing an attribute. So, it can be concluded the accuracy increase of 6.5659% in the implementation of the neuro fuzzy method and information gain in determination of landslide prone areas.

Keywords: Neuro Fuzzy, Information Gain, Fuzzy Logic, Artificial Neural Network, Landslide

1. INTRODUCTION

Nowadays, the accuracy of information is increasing rapidly as technological development. [1]. For the example, the information in determination of disaster severity. It can solve the problems which occur in society [2]. Then it makes a change in human life especially in determining disaster [3]. The disasters that can be determined is landslide. Landslide disasters affect people's life throughout the world, especially in developing countries such as Indonesia. Management of landslides draws attention by research from various fields of disciplines such as computer science, the science environment, meteorologists, geologists and scientists from various other disciplines who make joint efforts to determine the place and severity of landslides [4]. In the field of computer science, data mining aims to extract information from a data set so it can be read by users [5]. In addition, data mining can be applied by disciplines such as artificial neural networks, pattern recognition, spatial analysis of data, image databases and signal processing [6]. Artificial neural networks

can be used for disaster mitigation as disaster management, appropriate information flow, channeling assistance work and needs analysis on the concerns of victims. Artificial neural networks use information from the past to analyze the results of a particular problem or situation that might arise. Diverse data sources are collected regularly before, during, and after disasters by satellites, wireless and remote sensors, national meteorology and geology departments, government and private entities. The data collection has the requirements called 'Big Data' because of the volume, variation and speed at which data is produced [7]. The data will be processed with artificial neural networks, Takagi-Sugeno type (TS) neuro fuzzy for classification of landslide-affected populations, and one important input from the network is the output of another TS-type that repeats to recognize movement patterns from landslide location data. The experimental results show that neuro fuzzy shows good classification [8].

Neuro fuzzy is a combination of two systems, fuzzy logic system and artificial neural network. Neuro fuzzy systems are based on fuzzy inference systems that are trained using neural network learning algorithms [9]. Fuzzy Neuro has several advantages, one of them is to build a combination of fuzzy inference system (FIS) mechanisms which are described in neural network architecture. The form of FIS obtained from the clustering results are then changed using the order-one Sugeno fuzzy method [10].

Based on the results of a study Pradhan et al. [8] explained that the determination of landslide-prone areas in Penang Hill, Malaysia using the decision tree (DT), support vector machine (SVM) and neuro fuzzy results varied from 0.8204 to 0.9421 for success rate curves and 0.7580 to 0.8307 for each prediction level curve. In addition, the prediction curve shows that the decision tree predictive performance is slightly higher (83.07), while the success rate shows the neuro fuzzy method has a better prediction (94.21) capability among all of the methods used. Before processing data, data will enter the pre-processing stage. One of the processes applied is feature selection (FS). FS works by selecting features that have the potential to improve the accuracy of the classification algorithm [11]. The feature selection method used is information gain. After preprocessing the data, then the classification stage uses artificial neural networks.

Landslide is one of the most frequent and destructive natural disasters [12]. This disaster is caused by different physical conditions between one region and another, so that the trigger factors for landslides and material moving in the process are also different. One district that has the potential to experience landslides is Banjarnegara. It consists of at least 20 potential sub-districts. Part of the area is hilly with elevations ranging from 710 - 1,025 masl so that it is prone to landslides. The causes of landslides are eight landslide conditioning factors, namely distance to river, distance to road, height, slope, NDVI (Normalized Difference Vegetation Index), Band Ratio, SQBR (square root of band ratio) and vegetation index [13].

The application of information gain to data can improve the accuracy of the neuro fuzzy method. The purpose of this study is to apply information gain on the selection of landslide data features in determining landslide-prone areas with the neuro fuzzy method and to know the accuracy of implementing the neuro fuzzy method and neuro fuzzy method using information gain in determining landslide-prone areas.

2. METHODS

2.1. Neuro Fuzzy

According to Altaher et al. [14] Neuro fuzzy is a combination of two systems, fuzzy logic system and artificial neural network. Fuzzy logic is defined as a vague estimate which has a degree of membership ranging from 0 to 1 [15]. The neuro fuzzy system is based on fuzzy inference systems that are trained using learning algorithms derived from artificial neural network systems thus, the neuro fuzzy system has all the advantages possessed by fuzzy inference systems and artificial neural network systems. The neuro fuzzy structure is presented in Figure 1.
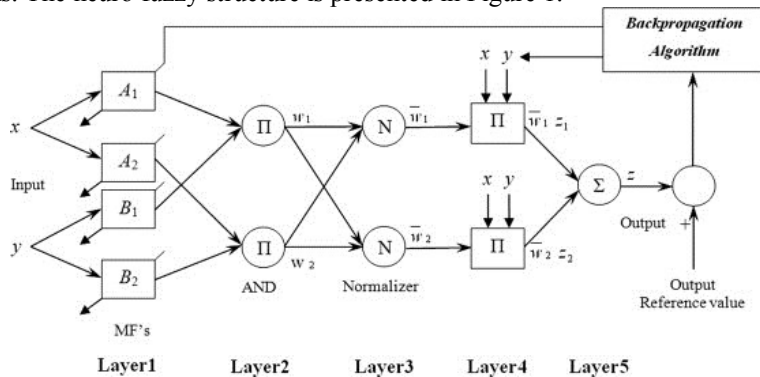

Figure 1. Neuro Fuzzy Structure

In the neuro fuzzy system there are five layers of processes in which the functions and equations of each layer are explained as follows:

Layer 1

All nodes in this layer are adaptive nodes (parameters can change). The parameters a, b, c in the gbell membership function are called the adaptive premise parameters, with b which is usually worth 1 to determine the bell membership value is not reversed. Here is a picture of the Generalized Bell curve which can be seen in Figure 2.
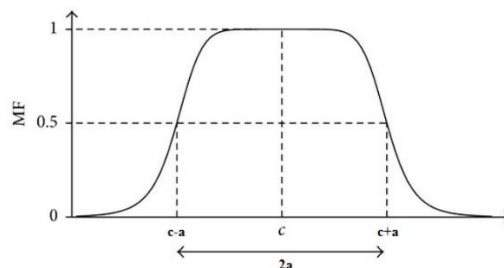

Figure 2. Generalized Bell Curve

The following gbell equation can be seen in Equation 1 as follows:

$$gbell(x, a, b, c) = \frac{1}{1 + \left|\frac{x - c}{a}\right|^{2b}} \qquad (1)$$

Description:
x = Input Value
a = the width of MF *(Membership Function)*
b = control *slopes* (coefficient values are usually valuable 1)
c = the center of MF *(Membership Function)*

Layer 2
All nodes in this layer are non-adaptive (fixed parameters). The function of this node is to multiply each incoming signal with the function of second layer node can be seen in Equation 2.

$$O_{2,i} = w_i = \mu_A(x) * \mu_B(y), i = 1,2 \qquad (2)$$

Description:
x = Input Value
y = Input Value
w_i = The Weight to-i
μ_i = Membership class value-i
$O_{2,i}$ = Result Value of Second Layer-i

Each node output represents the firing strength of each fuzzy rule. This function can be expanded if the premise has more than two fuzzy sets. The number of vertices in this layer indicates the number of rules that are formed.

Layer 3
Each node in this layer is a non-adaptive node that displays the normalized firing strength function, which is the ratio of the output of the i-node in the previous layer to all previous layer outputs, with the form of third layer node function can be seen in Equation 3.

$$O_{3,i} = \bar{w} = \frac{w_i}{w_1 + w_2}, i=1,2 \qquad (3)$$

Description:
wi = The Weight to-i
$\bar{w}i$ = Normalized Firing Strength to-i
$O_{3,i}$ = Result Value of Third Layer to-i

Layer 4
Each node in this layer is an adaptive node with the fourth layer node function seen in Equation 4.

$$O_{4,i} = \overline{w_i} \, f_i = \overline{w_i} \, (p_i x + q_i y + r) \qquad (4)$$

Description:
$\overline{w_i}$ = Normalized Firing Strength to-i
f_i = Control Signal-i
$O_{4,i}$ = Reasult Value of Fouth Layer to-i

By $\overline{w_l}$ is the normalized activation level of third layer and the parameter p, q, r states the consequent adaptive parameters.

Layer 5
In this layer there is only one fixed node whose function is adding all inputs, with the function of fifth layer node can be seen in Equation 5.

$$O_{5,i} = \sum_i \overline{w_l} f_i \qquad (5)$$

2.2. Neuro Fuzzy Learning
In the neuro fuzzy structure of Figure 1, the adaptive node is found in the layer first and fourth. The node in the first layer contains a parameter the premise is nonlinear whereas in the fourth layer it contains linear consequent parameters. [16]. To update these parameters it is necessary to use a method or algorithm. Neuro fuzzy uses a hybrid algorithm with two learning methods consisting of two parts namely forward pass and backward pass. In the forward direction, the premise parameter is fixed. By using the Recursive Least Square Estimator (RLSE) method, the consequent parameters are fixed based on the input-output data pair. The RLSE method can be applied because the consequent parameters that are corrected are linear parameters. The RLSE method will speed up the hybrid learning process. Then after the consequent parameters are obtained, the input data is passed back to the adaptive network and this adaptive network output is compared to the actual output. In the reverse direction, the consequent parameters are fixed. Errors that occur between adaptive network output and actual output are propagated back using gradient descent to improve premise parameters. This learning is known as the backpropagation-error algorithm. One step in the direction of back and forth learning is called one epoch.
The neuro fuzzy hybrid algorithm process can be seen in Table 1.

Table 1. The Neuro Fuzzy Hybrid Algorithm Process

	Forward Pass	Back Pass
Premise Parameter	Constant	Backpropagation Error
Consistent Parameters	LSE	Constant

2.3. Feature Selection
Feature Selection (FS) is a search process or technique used for feature selection to build strong learning models, such as artificial neural networks and decision trees. Feature selection can be used for the classification process [17]. The choice of features used for classification can have an impact on the accuracy of the classification function, the time needed for classification, the requirements for managing training data, and the implementation costs associated with classification.

2.3.1. Information Gain

The implementation of feature selection is conducted to get the gain value of each feature. This can be used as a basis for doing so to reduce it [18]. Information gain of an attribute, obtained from the entropy value before separation minus entropy after separation. This is the process of calculating information gain:

a. Find the value of entropy before separation by Equation 6 as follows:

$$Entropy\ (S) = \sum_{i-1}^{k} (P_i)\ log2(P_i) \tag{6}$$

b. Find the value of entropy after seperation by Equation 7 as follows:

$$Entropy\ (S,A) = \sum_{i-1}^{k} \left(\frac{|Sv|}{s} * Entropy(Sv) \right) \tag{7}$$

With v is all possible values of attribute A, and Sv is a subset of S where attribute A is worth v.

c. Find the value of information gain by Equation 8 as follows:

$$Gain\ (S,A) = Entropy(S) - Entropy\ (S,A) \tag{8}$$

The amount of information gain indicates how much an attribute affects the classification of data.

3. RESULTS AND DISCUSSION

In this research, the implementation of the method was conducted by using MATLAB R2017b tools. The data numbered 274 data related to landslides including rainfall, slope, soil type, geological structure, land use, hydrogeology and erosion prone. The type of data used is secondary data. These data are generally data on landslide-prone areas in 2011-2020 in Banjarnegara District, Central Java, Indonesia. Data is obtained from BPS (Central Statistics Agency), BPBN (National Disaster Management Agency), and the Agriculture Departement of Banjarnegara Regency. The data obtained has the .shp format then converted into excel form with ArcView GIS 3.3 tools. Table 2 describes the attributes of the following landslide data.

Table 2. Attributes of The Following Landslide Data

No.	Attribute	Data Type	Data Source	Description
1	Rainfall	Numeric	BPS	Rainfall (mm/year)
2	Slope	Numeric	BPBN	Slope Level (^{o})
3	Soil Type	Numeric	Agriculture Departement	Latosol Soil Type (Hectare)
4	Geological Structure	Numeric	BPBN	Geological structure of Miocene types of sedimentary facies (Hectare)
5	Land Use	Numeric	Agriculture Departement	Land use for settlements (Hectare)
6	Hydrogeology	Numeric	Agriculture	Hydrogeology with

			Departement	productive aquifers (Hectare)
7	Erosion	Numeric	BPBN	Areas prone to erosion (Hectare)
8	*Class*	Nominal	BPBN	No Prone dan Prone

Before the classification process uses the neuro fuzzy method, the data must be processed or commonly known as pre-processing. The purpose of pre-processing is to minimize errors and optimize mining results from the classifier used. The pre-processing used in this study consists of 2 stages.

3.1. Data Transformation Stage
The data transformation stage is conducted by transforming the value of nominal features into numeric types. Because its landslide data can be processed and classified in the form of applications made using Matlab tools. Each value of the nominal feature is transformed into 0 and 1. In the landslide class has 2 values. The transformation of the data is shown in Table 3.

Tabel 3. Result Transformation

Values Class	Result Transformation
Prone	1
No Prone	0

3.2. Feature Selection Stage
The feature selection stage has a dimension reduction in the data to optimize attributes that will affect the accuracy of the neuro fuzzy method. The dimension reduction in this configuration is conducted by using information gain techniques. The elimination of attributes is done one by one from the attributes that have the smallest information gain value and will then be mining. This disposal and mining process will stop when the accuracy results decrease. The following results of the gain value of each feature in the landslide data after being sorted from the highest to the lowest gain is shown in Table 4 as follows.

Table 4. The Gain Value of Each Attributes

No.	Attributes	Gain Value
1.	Erosion	0.5970
2.	Slope	0.2944
3.	Rainfall	0.2101
4.	Geological Structure	0.2044
5.	Soil Type	0.1310
6.	Land Use	0.1254
7.	Hydrogeology	0.0940

3.3. Classification Stage
The design of this application is implemented in the form of a GUI on Matlab. As for the classification phase, the data will be loaded into the application, then there is the Information Gain function button to generate the gain value of the feature and the Accuracy Result button to obtain the accuracy of the method.

The classification is carried out in 2 processes. First, the implementation the classification of landslide data with the neuro-fuuzy method. Second, the implementation of classification with the neuro-fuuzy method and information gain. In this study, the distribution of training data and testing data uses 20 k-fold cross validation. The data will be divided into 20 sub-data / sub-samples and the data will also be 20 as iterations. It means every sub-sample will be carried out a classification process to get its accuracy.

3.3.1. The Classification Using Neuro Fuzzy

The classification process is determining input variables and output variables. Input variables are factors that cause landslides, namely rainfall, slope, soil type, geological structure, land use, hydrogeology and erosion prone. Output variables are vulnerable and not prone to landslides. The results of the accuracy obtained is shown in Table 5.

Table 5. Result Accuracy of Neuro Fuzzy

Method	Accuracy
Neuro Fuzzy	81.9231%

3.3.2. The Classification Using Neuro Fuzzy and *Information Gain*

The classification process with the neuro fuzzy method starts to reduce from the attribute which has the smallest gain value and it will stop when the accuracy is increased and there is no decreasing accuracy. The attribute reducing process with the neuro fuzzy method by applying information gain as a feature selection is presented in Table 6.

Table 6. The Attribute Reducing Process

K-fold	Attribute Reducing
1	1
2	1
3	1
4	1
5	2
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	2

The accuracy of implementing neuro fuzzy and information gain is shown in Table 7.

Table 7. Result Accuracy of Neuro Fuzzy and Information Gain

Method	Accuracy
Neuro Fuzzy and Information Gain	88.489%

Based on the classification process, the removal is carried out in just 1 attribute. It showed that the highest accuracy is obtained by applying information gain is 88.489%. The reduced attribute is Hydrogeology.

4. CONCLUSION

The implementation of the neuro fuzzy method on landslide data was obtained an accuracy of 81.9231%. In the implementation of the neuro fuzzy method with information gain was conducted in classification process. The process will stop when the accuracy has decreased. The highest accuracy result was obtained of 88.489% by removing an attribute. So, it can be concluded the accuracy increase of 6.5659% in the implementation of the neuro fuzzy method and information gain in determination of landslide prone areas. The graph of increasing in the neuro fuzzy accuracy can be seen in Figure 3.
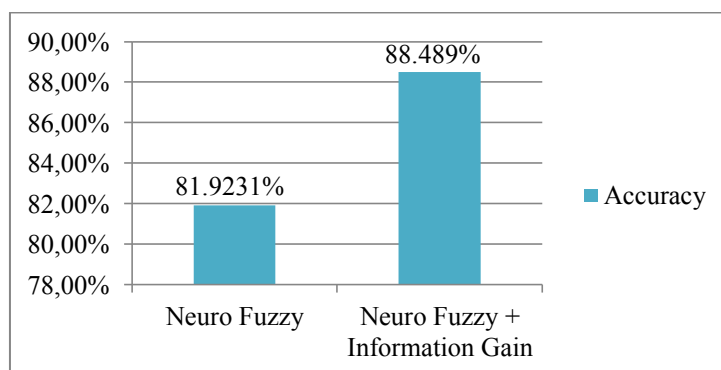

Figure 3. Graph of Increasing Neuro fuzzy and Neuro fuzzy + Information Gain Accuracy

5. REFERENCES

[1] Muslim, M. A., Prasetiyo, B., & Alamsyah. (2016). Implementation Twofish Algorithm for Data Security in A Communication Network Using Library Chilkat Encryption Activex. *Journal of Theoretical and Applied Information Technology*, 84(3), 370-375.

[2] Pramesti, A. A., Arifudim, R., & Sugiharti, E. (2016). Expert System for Determination of Type Lenses Glasses using Forward Chaining Method. *Scientific Journal of Informatics*, 3(2), 177- 188.

[3] Muslim, M. A. & Retno, N. A. (2014). Implementasi Cloud Computing Menggunakan Metode Pengembangan Sistem Agile. *Scientific Journal of Informatics*, 1(1), 29-38.

[4] Hristidis, H., Chen, S. H., Li, T., Luis, S. & Deng, Y. (2010). Survey of Data Management and Analysis in Disaster Situations. *The Journal of Systems and Software*, 83(1), 1071-1714.

[5] Refonaa, J., Lakshimi, M., & Vivek, V. (2015). Analysis and Prediction of Natural Disaster Using Spatial Data Mining Technique. *Proceeding of International Conference on Circuit, Power and Computing Technologies*, 37 (2), 1-6.

[6] Sugiharti, E. & Muslim, M. A. (2016). On-Line Clustering Of Lecturers Performance Of Computer Science Department Of Semarang State University Using K-Means Algorithm. *Journal of Theoretical and Applied Information Technology*, 83(1), 64-71.

[7] Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A. & Chakraborty, B. (2016). A Review on Application of Data Mining Techniques to Combat Natural Disasters. *Ain Shams Engineering Journal*, 33(13), 1-14.

[8] Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, 51(1), 350-365.

[9] Tavakkoli, P., Souran, D.M., Tavakkoli, S., Hatamian, M., Mehrabian, A. and Balas, V.E. (2015). Classification of The Liver Disorders Data Using Multi-Layer Adaptive Neuro-Fuzzy Inference System. *Proceeding of In Computing, Communication and Networking Technologies,* 1-4

[10] Akcay, O. (2015). Landslide Fissure Inference Assessment by ANFIS and Logistic Regression Using UAS-Based Photogrammetry. *ISPRS International Journal Geo-Informatio*, 4(1), 2131-2158.

[11] Ahmad, F., Isa, N.A.M., Hussain, Z., Osman, M.K. and Sulaiman, S.N. (2015). A GA-Based Feature Selection and Parameter Optimization of An ANN in Diagnosing Breast Cancer. *Pattern Analysis and Applications*, 18(4), 861-870.

[12] Chen, W., Panahi, M. and Pourghasemi, H.R. (2017). Performance Evaluation Of GIS-Based New Ensemble Data Mining Techniques of Adaptive Neuro-Fuzzy Inference System (ANFIS) With Genetic Algorithm (GA), Differential Evolution (DE), And Particle Swarm Optimization (PSO) for Landslide Spatial Modelling. *Catena*, 157(2017), 310-324.

[13] Wan, S., Lei, T. C., & Chou, T. Y. (2010). A Novel Data Mining Technique of Analysis and Classification for Landslide Problems. *Natural Hazards,* 52(1), 211-230.

[14] Altaher, A., Almomani, A. & Ramandas, S. (2012). Application of Adaptive Neuro-Fuzzy Inference System for Information Secuirty. *Journal of Computer Science*, 8(6), 983-986.

[15] Pamuji, A. (2017). Fuzzy Logic Inference System for Determining the Quality Assesment of Student's Learning ICT. *Scientific Journal of Informatics*, 4(1), 57-65.

[16] Walia, N., Singh, H. & Sharma, A. (2015). ANFIS: Adaptive Neuro-Fuzzy Inference System-A Survey. *International Journal of Computer Applications*, 123(13), 32-38.

[17] Kabir, M. M., Islam, M. M. & Murase, K. (2010). A New Wrapper Feature Selection Approach Using Neural Network. *Neurocomputing,* 73(16-18), 3273-3283.

[18] Muslim, M. A., Nurzahputra, A., & Prasetiyo, B. (2018). Improving Accuracy of C4. 5 Algorithm Using Split Feature Reduction Model and Bagging Ensemble for Credit Card Risk Prediction. *Proceeding of International Conference on Information and Communications Technology*, 1(2018), 141-145.