Artificial neural network for predicting earthquake casualties and damages in Indonesia

Rienna Oktarina¹, Senator Nur Bahagia², Lucia Diawati², Krisha S. Pribadi³

¹ Industrial Engineering Department, Faculty of Engineering, Bina Nusantara University Jakarta, Indonesia 11480
² Industrial Engineering Department, Bandung Institute of Technology, Bandung, Indonesia 40132
³ Civil Engineering Department, Bandung Institute of Technology, Bandung, Indonesia 40132

Email: rienna.oktarina@binus.ac.id

Abstract. The paper is intended to develop a model to predict the number of damaged buildings and casualties due to earthquake using ANN (Artificial Neural Network). This model is expected to be able to generate the type and amount of relief supplies required by those affected during the emergency phase. This research develops ANN using supervised learning paradigm, and backpropagation learning algorithm. The applied ANN network architecture is a multiple-layer system, with 1 (one) neuron used in both input and output layer, and 95 (ninety-five) neurons used in the hidden layer yielding 0.99971 as the greatest value of the correlation coefficient. The output variable in this study is the earthquake impact consisting of six variables. While the input variables (predictors) in this study consisting of eight variables. The model in this study utilizes 123 seismic datasets, divided into 100 data (80%) for the training process and 23 data (20%) for the testing process. This research adds to the existing research and demonstrates the application of ANN in predicting the numbers of damaged buildings and casualties. The model is useful in supporting and strengthening preparedness and emergency relief activities due to earthquake disaster.

Keywords: Artificial Neural Network, Earthquake damage prediction, Earthquake casualties prediction

1. Introduction

Indonesia is geographically situated at the intersection of four tectonic plates, i.e. Asian plate, Australian plate, Indian Ocean plate, and Pacific Ocean plate and moreover, it also lies on the “ring of fire” zone, contributing to higher disaster risks, one of which is earthquake disaster. Supported by compiled data from The International Disaster Database (http://www.emdat.be/database), Indonesia had experienced at least 17 earthquake occurrences, causing 1,431 people died and 2,933,205 people were affected for the last six years. Earthquake causes serious disruption to the functioning of society, causing widespread loss of human life in terms of material, socio-economic, causing deaths and injuries, and causes physical or environmental damage [1]. Figure 1 defines the characteristics and consequences of disasters as stated by Chen [1].
Considering the earthquake-generated disaster impacts to the affected society, it is necessary to undertake an assessment process on the occurred losses and damages as a contingency planning guidance for the responsible stakeholders. One example is to identify the numbers of collapsed houses in order to determine the numbers of tents required to accommodate the homeless. Another example is to assess the numbers of injured victims following earthquake disaster in order to manage the hospital capacity nearby, or an assessment of displaced persons to identify the amount of aid goods that must be provided to lessen their suffering. Therefore, development of a model in order to predict the losses and damages due to earthquake disaster has become essential.

Previous researchers have worked on the loss and damage prediction theme, such as Badal et al.[2] who have developed a model to predict the approximate number of casualties and cost loss associated with damage in the event of an earthquake disaster in urban areas. There are three important stages in the creation of this model, i.e.: (1) estimating the epicentral intensity, (2) selecting the attenuation law, and (3) calculating the number of victims. Other previous works focus on loss prediction, in particular economic loss prediction, have been done by Yong et al. [3], Yong et al. [4], Ergonul [5], Chen et al. [6], and Chongfu [7].

Other previous researchers have focused on the prediction of earthquake problems using ANN is Asim et al.[8] and Asim et al.[9] who carried out Earthquake Magnitude Prediction in Hindukush Region. González et al.[10] also conducted earthquakes magnitude prediction using Neural Networks. Whereas Perol et al.[11] have conducted earthquake detection and location using ANN, and Saba et al.[12] conducted based earthquake prediction for Pakistan region. Thus, it can be concluded that ANN can be used for predicting earthquake casualties and damages in Indonesia.

### 2. Model Development

According to the explanation on the previous section, it is summarized that prediction on losses and damages following an earthquake occurrence is deemed important. Therefore, this current research will develop a model to predict the number of damaged buildings, casualties, fatalities and displaced people caused by the earthquake, using Artificial Neural Network (ANN). According to Fausett [13], ANN is a model to match or imitate biological neural networks which nodes are based on simplified mathematical representations of real neurons. ANN is used in this research as it possesses an immunity or high level of tolerance, so that ANN keeps functional and is still able to yield complete output regardless incomplete input data, thanks to its manipulating capacity. This is appropriate for
Indonesian case considering the lack of complete and detailed information on disaster occurrences database in the country.

Model development on this research consist of four stage, are: (1) identification phase and data collection, (2) neural network design phase, (3) development of neural network stage, and (4) analysis and conclusions. Figure 2 illustrates the four stages of model development on this research. Detailed explanation on the process and result of each stage are as follows.

2.1. Identification Phase and Data Collection

2.1.1. Input and output variables
The input variables (predictors) are influential in identifying the target (output) while the output variables as the target are dependent and need to be identified. In this research, output variable or predicted variable is the impacts of earthquake disasters. According to Chen [1] on Figure 1, the earthquake impacts vary from physical damages, such as collapsed buildings houses or death and injuries. Therefore, for this research, the output variables are related to the earthquake impacts, which consist of six variables, they are: (a) number of fatalities; (b) number of casualties; (c) number of Internally Displaced Persons (IDPs); (d) number of severely damaged houses; (e) number of moderately damaged houses; and (f) number of least damaged houses. Whereas input variables used in this research are the earthquake parameters used in RADIUS programme. The input variables (predictors) in this research consist of eight variables, they are: (a) total population; (b) area; (c) population density; (d) population index; (e) time of earthquake occurrence, whether it occurs on day time (6:00 to 18:00 hours) or night time (6:00 p.m.-6:00 hours); (9) earthquake magnitude; (g) earthquake depth, and (h) distance to the earthquake epicentrum.

2.1.2. Data collection
The historical data tracking regarding the earthquake incidence occurred in Indonesia becomes the basis for data collection. Data on the earthquake impacts (variable output) are well documented in the DIBI (Data & Information Disaster in Indonesia) database, compiled by BNPB (National Disaster Management Agency), and accessible via http://dibi.bnpb.go.id with recorded data on disaster impacts.
from year 2000 to 2019. DIBI Database provides 456 earthquake occurrences data, which narrowed down to 123 qualified data following a complete cross-section assessment. For the purpose of validation process, the data are divided into two groups (split-sample), one is data for training and the other is for testing. Training data uses approximately 80% randomly selected data to build the model while the remaining 20% of the overall data is used for validation process on data testing.

2.2. Artificial Neural Network Design

2.2.1. Network architecture

According to Bal and Buyele-Bodin [14] and Cachim [15], the most widely used ANN type is the multi-layer perceptron with backpropagation training algorithm for minimization of error. Therefore, ANN network architecture used in this research is a multiple-layer system, which consists of an input layer, hidden layer and output layer. The process of determining the number of neurons in the hidden layer includes trial and error and consideration on the value of correlation coefficient. Correlation coefficient value close to one indicates a good match between the levels of output and the target output. The number of neurons used in the input layer and output layer is 1 (one), while the number of neurons in the hidden layer is 95 (ninety-five) neurons, because this combination yields the best correlation coefficient value of 0.99971, close to 1.

Figure 3 schematically presents the ANN network architecture in this research, using 1000 times iteration of training data vectors (epoch) as the maximum value. It usually takes a lot epoch in backpropagation ANN learning process.

2.2.2. Learning methods selection

Two main processing phases of neural network include training and testing. The training process comprises the adjustment of weights and biases in order to obtain output by applying a proper method [16]. According to Dantas et al. [17], the supervised methods are the most general methods for training. The supervised methods use an algorithm, the least mean square method and its simplification to multilayer networks, which is the backpropagation algorithm. This research applies supervised learning as the paradigm and backpropagation as the algorithm.

2.2.3. The training process

Faussett [13] emphasizes that it is essential not to overtrain ANN. Assuming that more trainings will lead to better output is completely wrong. When ANN repeatedly receives same patterned trainings, the weights will be set closer to the expected output, resulting ANN to forcedly memorize the patterns (memorization), thus does not consider the existing relations (generalization). As ANN training process intends to balance the ability of memorizing and generalizing, applying untrained patterns will lead to unsatisfactory result. In brief, ANN is expected to be not only better at predicting sets of training data, but also providing good results when predicting the testing data sets and validating
The number of training data used in this research is 100 data sets (80% of 123 datasets) while testing data consists of 23 data sets (20% of 123 datasets).

3. Results & Interpretation

Upon completion of the training process, this research compares the actual data to the network output resulted by the training data. The best results occur when the output is equal to the actual data. Figures 4a - 4f present the comparison between the actual data and network output for the training data. The comparison consist of the number of fatalities, number of casualties, number of IDPs, number of severely damaged houses, number of moderately damaged houses, and number of least damaged houses respectively.
Figures 5a – 5f show the comparison between actual data and network output for the testing data. The comparison consist of respectively number of fatalities, number of casualties, number of IDPs, number of severely damaged houses, number of moderately damaged houses, and number of least damaged houses.

Figure 5a. Comparison Number of Fatalities between Target and Output for Testing Data

Figure 5b. Comparison Number of Casualties between Target and Output for Testing Data

Figure 5c. Comparison Number of IDPs between Target and Output for Testing Data

Figure 5d. Comparison Number of Severely Damaged Houses between Target and Output for Testing Data

Figure 5e. Comparison Number of Moderately Damaged Houses between Target and Output for Testing Data

Figure 5f. Comparison Number of Least Damaged Houses between Target and Output for Testing Data
By looking at the comparison between the target and network output for the training and testing data, it is clear that the network output intersect with the target. The best results occur when the results of network output is equal to the target, indicated by intersected graph lines. Table 1 below provides the Mean Square Errors (MSEs) values between the network outputs and the actual data for the training and testing data.

Table 1 shows that the MSEs of several output variables on the testing data are much higher compared to those of the training data. It indicates large differences exist between the network outputs and actual data from the testing data sets. The designed ANN has been able to recognize the pattern of training data, however it has relatively low accuracy for prediction.

### 4. Verification and validation

Verification process ensures that the system is mechanically acceptable, while validation works on proving that a system or network has been developed correctly. Holdout technique used in this research validates large datasets, which are divided into training and testing datasets.

Finally, the development of this model involves the ANN model testing and performance model checking. Side by side with the output of RADIUS Programme Darpito [18], Okazaki [19] and Badal [2], the outputs of the ANN model developed in this research compared to the actual data obtained from the earthquakes occurred in Yogyakarta and in Padang (Table 2). The model testing proves the validity of model. To verify the result of this research conducted by comparing the output of model with the actual data of earthquake occurrences in Yogyakarta in 2006 and Padang in 2009. Small error value of the model output proves the validity and in the case of large error value, further works are required to attain model validity. Table 2 compares the model output to the Badal et al [2], RADIUS programme [22] and actual data.

Table 2 shows the MSE value for the output of RADIUS program is 7,824,074; MSE for the output of Badal model is 43,024,790, and the MSE for the output model of this research is 520,924. MSE
value of the output model developed in this research is the smallest value compared to the output generated by the RADIUS programmed and Badal model, it indicates that the developed model is preferable than RADIUS programmed and model of Badal model.

5. Analysis and Conclusions
Table 2 illustrates the performance comparison among the ANN model, RADIUS model, and Badal model. According to Table 2, the prediction value produced by ANN model is close to actual data value. For example, ANN model output for prediction on collapsed/damaged houses in Klaten regency due to Yogyakarta earthquake is 92,529 units, compared to actual data of 92,967 units, showing differences of 438 houses or 0.47% error. The output of ANN model provides better output value as it is closer to the actual data.

Input variables used in this study is the earthquake parameters that can be immediately known while the earthquake occurred. Determining the input variables are important in designing ANN, because the accuracy of the type of input variables can affect the effectiveness or duration of the training process. The developed model is useful to predicting the number of damaged buildings and casualties due to earthquakes. Hence, the result accelerates the logistics assessment implementation, ensuring timely delivery of relief goods during the emergency response phase to help alleviate the suffering of the affected people. Apart from its higher predictive power, ANN model is also a robust model with less input data regardless its ability to predict more output, unlike the RADIUS programme. However, this developed ANN model still lacks of clear inter-relationship between the input variables and output variables.

References
[1] Chen L 2010 Optimization models and methodologies to support emergency preparedness and post-disaster response (Doctoral dissertation)
[2] Badal J, Vázquez-Prada M and González Á 2005 Preliminary quantitative assessment of earthquake casualties and damages Natural Hazards 34(3) 353-74
[3] Yong C, Qi-Fu C, and Ling C 2001 Vulnerability analysis in earthquake loss estimate Natural Hazards 23(2-3) 349-64
[4] Yong C, Ling C, Güendel F, Kulhánek O, and Juan L 2002 Seismic hazard and loss estimation for Central America Natural Hazards 25(2) 161-75
[5] Ergonul S 2005 A probabilistic approach for earthquake loss estimation Structural Safety 27(4) 309-21
[6] Chen Q F, Chen Y, Liu J I E, and Chen L 1997 Quick and approximate estimation of earthquake loss based on macroscopic index of exposure and population distribution Natural Hazards 15(2-3) 215-29
[7] Chongfu H 1996 Fuzzy risk assessment of urban natural hazards Fuzzy sets and systems 83(2) 271-82
[8] Asim K M, Martínez-Álvarez F, Basit A, and Iqbal T 2017 Earthquake magnitude prediction in Hindukush region using machine learning techniques Natural Hazards 85(1) 471-86
[9] Asim K M, Idris A, Iqbal T, and Martínez-Álvarez F 2018 Earthquake prediction model using support vector regressor and hybrid neural networks PloS One 13(7) e0199004
[10] González J, Yu W, and Telesca L Earthquakes magnitude prediction using recurrent neural networks
[11] Perol T, Gharbi M, and Denolle M 2018 Convolutional neural network for earthquake detection and location Science Advances 4(2) e1700578
[12] Saba S, Ahsan F, and Mohsin S 2017 BAT-ANN based earthquake prediction for Pakistan region Soft Computing 21(19) 5805-13
[13] Faussett L 1994 Fundamentals of neural networks: architectures, algorithms, and applications Prentice-Hall Inc
[14] Bal L, and Buyle-Bodin F 2013 Artificial neural network for predicting drying shrinkage of concrete Construction and Building Materials 38 248-54

[15] Cachim P B 2011 Using artificial neural networks for calculation of temperatures in timber under fire loading Construction and Building Materials 25(11) 4175-80

[16] Shafabakhsh G H, Ani O J, and Talebsafa M 2015 Artificial neural network modeling (ANN) for predicting rutting performance of nano-modified hot-mix asphalt mixtures containing steel slag aggregates Construction and Building Materials 85 136-43

[17] Dantas A T A, Leite M B, and de Jesus Nagahama K 2013 Prediction of compressive strength of concrete containing construction and demolition waste using artificial neural networks Construction and Building Materials 38 717-22

[18] Darpi K 2011 Earthquake damage assessment: Application and verification of the radius method to yogjakarta earthquake 2006 Bulletin of the International Institute of Seismology and Earthquake Engineering 45 85-90

[19] Okazaki K, Villacis C, Cardona C, Kaneko F, Shaw R, Sun J, and Tobin L T 2000 RADIUS: Risk assessment tools for diagnosis of urban areas against seismic disasters UN International Strategy for Disaster Reduction (ISDR) sd Geneva CH