

LOW DATA REQUIREMENTS FRAMEWORK FOR LANDSLIDE SUSCEPTIBILITY PREDICTION USING 3D ALOS PALSAR IMAGES AND NEURAL NETWORKS

Sohaib K. M. Abujayyab*, İsmail Rakıp Karas

* Dept. of Computer Engineering, Karabuk University, Demir Celik Campus, 78050 Karabuk, Turkey.
s.jayyab@hotmail.com and irkaras@gmail.com

KEY WORDS: ALOS PALSAR, Landslide Susceptibility, Neural Networks, Topographic Attributes, Low Requirements Framework.

ABSTRACT:

Development of landslides susceptibility (LS) predictors based on 3D data is an active area of research in the recent years. Predicting landslides susceptibility maps help to secure human lives and maintaining infrastructures from this risk. Several advanced frameworks proposed with high input data to improve the predictors. The aim of this paper is to develop low data requirement framework for LS predictors development. This framework is only using one input 3D ALOS PALSAR image. The framework has three stages. (A) data pre-processing, (B) deriving explanatory factors, and (C) neural networks training and testing. Exactly, 22 input spatial factors were extracted from ALOS PALSAR image. Extracted factors were utilized to develop the FFNN predictor. The structure of the predictor is 22 factors (input layer) \times 150 neurons (hidden layer) \times 1 (output layer). Furthermore, 5829 sample points utilized during the training stage, while 745810 points sent to the trained predictor to create LS map. Based on confusion matrix metric, performance accuracy (89.3% training and 82.3 testing), While (95.22% training and 84.7% testing) based on Receiver Operating Characteristic curve. Out of the study area in Karabuk, 3.53 km² (3.03%) were located in very high susceptibility category. Lastly, the application of the proposed framework showed that it is capable develop low data requirement predictors with high accuracy. Framework provide guideline data for future development in taxing topographic circumstances and large scale of data coverage. In addition, the framework handled the inconsistency in data quality and data updating problem.

1. INTRODUCTION

Landslides worldwide considered as a chief Geohazard to infrastructures, properties, natural environments, and lives in mountainous areas. Indicators from the statistics centres demonstration that landslides were fully accountable for 17% of all deaths from natural disaster globally (Chae et al. 2017). Landslides were responsible for thousands of materials damages and losses of billions. In turkey, this issue is anticipated to remain in the following years as a result of urbanization growth and continuous deforestation. In addition, the regional precipitation increment exposes the areas to landslides danger caused by fluctuating in climatic conditions (Yilmaz 2009). Landslides can be occurred through prolonged or heavy rainfall. Landslides can be happening through a group of geomorphic parameters, for instance land cover, geology, topography, forest, and soil. Therefore, in the last decades, prediction of landslides susceptibility (LS) maps become urgent issue between international geology, geomorphology and engineering communities (Nefeslioglu et al. 2012).

LS maps can be predicted by select relevant factors that have been contributed to historical landslides occurrence. Then, utilize relevant factors to develop the predictor and forecast the positions of upcoming landslides occurrence (Song et al. 2012). Forecasting future LS sites of assumed geographical zone needs a quantitative framework. Frameworks required to extract the sophisticated knowledge that affected previous landslides occurrence utilizing relevant parameters of landslides. In recent years, several approaches have been recommended to develop the predictor such as data mining, statistics, probability

analysis, as well as sophisticated approaches like Neural Network (NN) with an parallel employment of remote sensing and geographic information system (Chae et al. 2017; Chaudhary et al. 2015; Conforti et al. 2014; Felicísimo et al. 2013; Li, Wang, and Li 2015; Shahabi et al. 2014).

In recent times, NN has been an developed method for several research areas (Astray et al. 2016; Choi et al. 2012; Fu et al. 2013; Sener et al. 2012; Song, Zheng, and Li 2010; Yeo and Yee 2014; Yilmaz 2009). Main advantages of utilizing NN that it is capable to work with any type of data at any measurement scale for ordinal, nominal, ratio and linear distribution. NN are data-driven models and universal non-linear function for prediction. Furthermore, NN can effectually treat spatial data and gain high performance accuracy during the predictor development and deployment (Yilmaz 2009).

Although LS predictor is highly influenced by employed methods, the quantity and quality of the input data have a similar effect on outcomes. In the case of performing LS prediction at large-scaled, considerable consumed time required to gather well-documented topographical data, geological data and landslide records. In the event of working in mountainous zones, to collecting relevant data of landslides occurrences and other parameters from the field are challenging due to the difficult mountainous conditions (Nefeslioglu et al. 2012). To mitigate the challenges, 3D satellite images have been widely utilized. Several 3D satellite images were available for research analysis, for instance; SRTM 30m (Shuttle Radar Topography Mission), ASTER 30m (Advanced Spaceborne Thermal Emission and Reflection Radiometer), and ALOS PALSAR

* Corresponding author

12.5m (Advanced Land Observing Satellite - Phased Array type L-band Synthetic Aperture Radar) (Japanese Earth observing satellite 2019). 3D Satellite images can be employed to automatically extract several relevant parameters for LS predictors development.

The main objective of this study is to introduce low data requirement framework for LS predictors development. The framework mainly relies on 3D ALOS PALSAR images and prediction process performed using neural networks. This paper consisting 4 section, first the introduction, methodology (consisting study area, data collection, proposed framework, accuracy evaluation metrics and predictor establishment), then result analysis and discussion. Lastly the conclusion.

2. METHODOLOGY

2.1 Study area

The study area located in the northwest of Turkey in Karabuk state. Geographically it lies between range from latitude $40^{\circ}58'30''\text{N}$ to $41^{\circ}07'40''\text{N}$, and longitude $32^{\circ}43'10.09''\text{E}$ to $32^{\circ}53'15.53''\text{E}$ (Figure 1). The entire area approximately 153 km^2 , which account around 3.7% of the complete area of Karabuk state. In 2018, entire population of Karabuk state was 250,269 persons. Climate is classified as warm and temperate. Karabuk has a significant amount of rainfall during the year. The average temperature is 9.9°C . Around 733 mm of precipitation falls annually. The study area is sub river watershed. The study area is totally mountains area. Average of elevation is 1024 meters above mean sea level. The highest elevation is 1590 meters. Consequently, several past landslides happened in this area. The study area characteristic is very suitable to apply LS framework.

2.2 Data collection

In order to produce the LS map and perform NN predictions, relevant factors must be determined. In next section, all

parameters will be extract from digital elevation model DEM. DEM is a digital geographic values in three dimensions (3D) (x, y and z coordinates), where z represent the altitude value above the mean sea level.

In this study, DEM collected. DEM obtained from Advanced Land Observing Satellite (ALOS) platform. ALOS was a mission of the Japan Aerospace Exploration Agency (JAXA) and Japanese earth observing satellite program (Alaska Satellite Facility's 2011). ALOS platform has three remote-sensing instruments. PALSAR (Phased Array type L-band Synthetic Aperture Radar) is one of the instruments that provide DEM images. PALSAR developed to contribute to the fields of resource surveying, precise regional land-coverage observation, mapping, and disaster monitoring. PALSAR is an active microwave sensor using L-band frequency to achieve cloud-free and day-and-night land observation (Japanese Earth observing satellite 2019).

The temporal extent is 2006-05-16 to 2011-04-21, while the repeats coverage is 46 days. Spatially PALSAR images have global coverage and provided under open access policy. PALSAR DEM images have spatial resloluation at 12.5 m, which quite good for LS prediction. This high spatial resolution helps the predictors to achieve high performance accuracy especially in case of small scale landslide occurrence. In addition, high resolution can replace insufficiency arising from limited input data or excluding modest accuracy input data for instance modest accuracy of soil map. The images consisting 16 Bit of pixel depth, which is richer than past DEM. Rich pixel DEM better represent the real elevation especially in complicated areas like mountains areas. Additionally, this research will extract all predication parameters only from the DEM, So the input accuracy will affect and control all the data in the predictor. Lastly, past landslides locations were collected and mapped as illustrated in Figure 1.

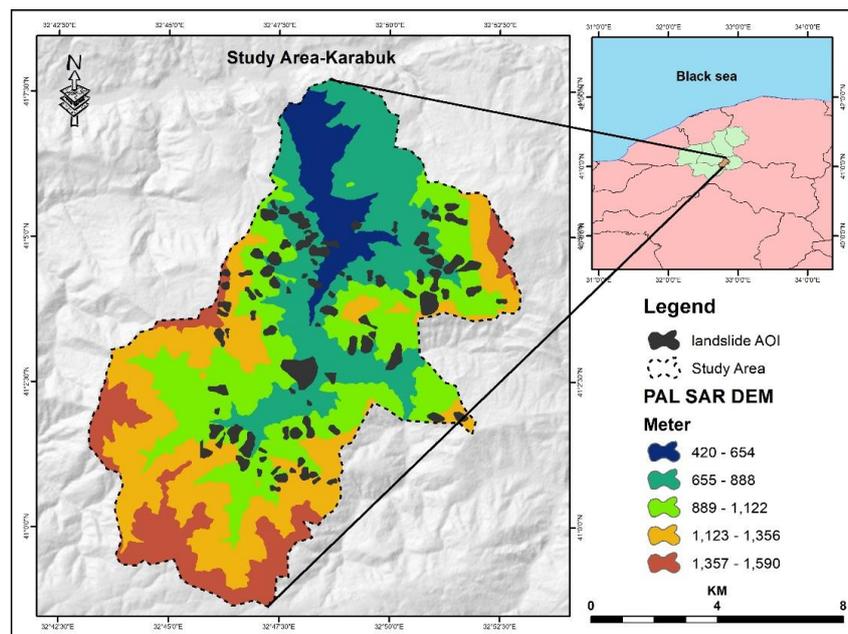


Figure 1 Study area, PAL SAR Digital elevation predictor and past landslides locations.

2.3 Low data requirement framework

This section is introducing low data requirement framework for LS predictor development (Figure 2). The framework mostly founded based on ALOS PALSAR DEM images. The framework consisting three stages; (A) the first stage is data pre-processing. DEM were collected and feed to ArcGIS environment. Then data restricted to study area by implementing Mask tool. In order to handle gaps problem in case it is existing, Fill tool utilized to treat the input image. (B) the second stage is mainly about deriving explanatory factors from the DEM image. In this study, 21 factors were extracted from the DEM image. Derived explanatory factors represent the surface topographic, Geomorphology texture, Geomorphometry, temperature and moisture factors Table 1.

The last stage (C) is about employing neural networks to perform LS training. Then use the trained network to produce LS maps. In this stage NN training dataset required to be construct by applying sampling. Then extract samples data from the derived factors images and target image (past landslides locations).

Table 1 List of explanatory factors.

N.	Group	Factors
1	Topographic surfaces and Geomorphology texture	1. Processed Elevation 2. General Curvature 3. Plan Curvature 4. Profile Curvature 5. Dissection 6. Roughness 7. Landform 8. Slope Position 9. Surface Area Ratio 10. Surface Relief Ratio 11. Aspect 12. Hillshade 13. Slope
2	Geomorphometry	14. Flow Direction 15. Flow Accumulation 16. Stream Order
3	Temperature and moisture	17. Heat Load Index 18. Integrated Moisture Index 19. Site Exposure Index 20. Compound Topographic Index 21. Slope/Aspect Transformation 22. 2nd Derivative Slope

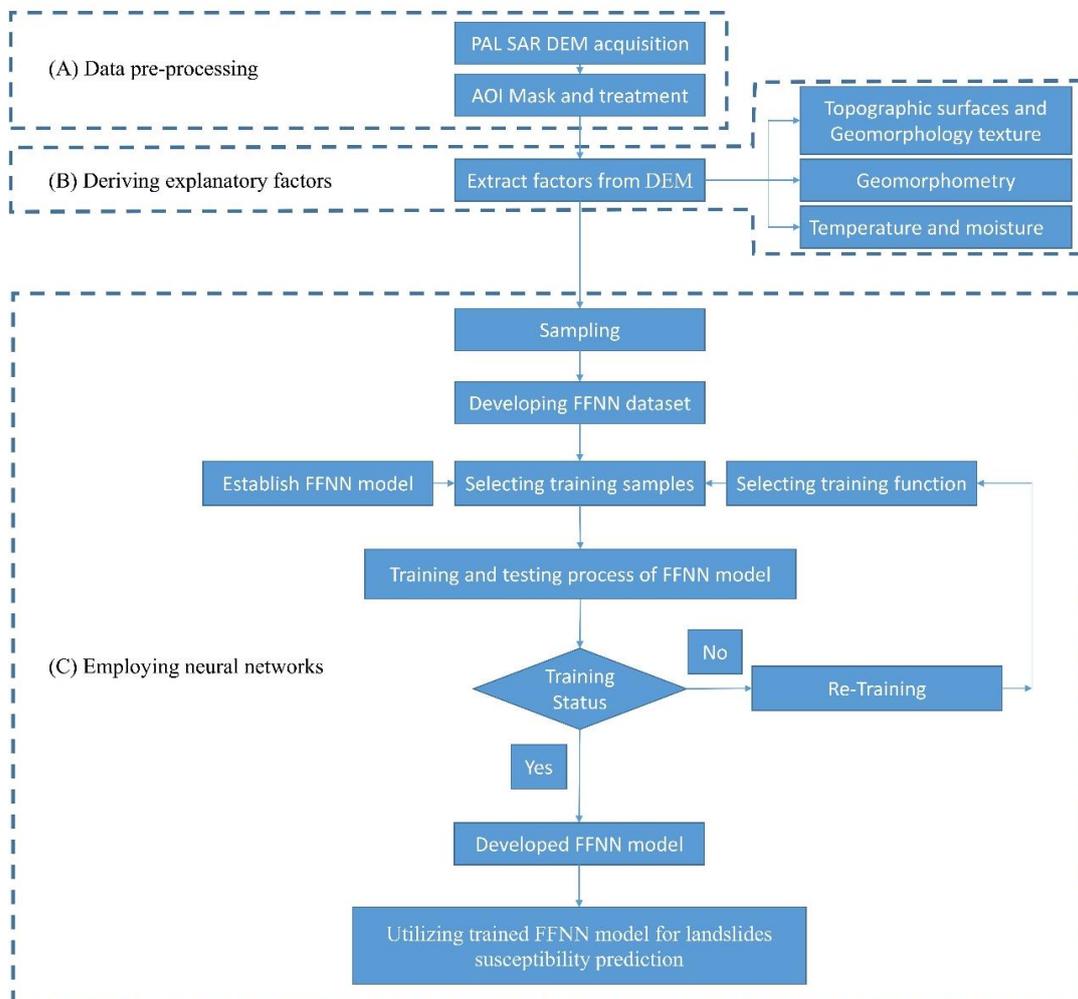


Figure 2 Low data requirement framework for LS prediction.

2.4 Accuracy evaluation metrics

The chief consideration of the NN predictor in this study is the performance accuracy. Accuracy of NN acquired by the trade-off between actual and calculated output. The clarification of accuracy illustrate throughout confusion matrix CM as clarified in equations (1) (Liang et al. 2015). The performance accuracy calculated for both training and testing datasets.

$$CMA = 1 - \left(\frac{\text{Number of samples correctly classified}}{\text{Total number of samples}} \right) * 100 \quad (1)$$

Where;

CMA = confusion matrix Accuracy,
TP = true positive,
TN = true negative,
FP = false positive, and
FN = false negative

In addition, (B) for additional conformation of performance accuracy, metric of Receiver Operating Characteristic (ROC) curves extracted. The ROC metric define the performance accuracy through area under curve (AUC or AUROC) (Conforti et al. 2014; García-Rodríguez and Malpica 2010).

2.5 LS predictor establishment based on NN

The feed forward neural network (FFNN) utilized in this paper is a three-layered network learning by the generalized delta rule. Before running the FFNN program, the training sites were selected. In this paper, 5829 training points representing landslide past zones and non-landslide zones were used. Samples were randomly defined in ArcGIS software. In order to construct and train the FFNN predictor, samples dataset migrated to Matlab software. Based on, spatial data of landslides in Karabuk area, the three layered (Input, hidden, and output) FFNN predictor developed.

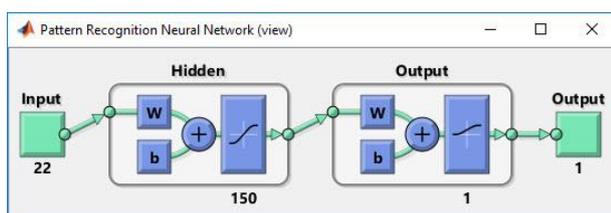


Figure 3 Structure of FFNN predictor for LS.

150 neurons were defined for the hidden layer after several experiments from 10 to 200 neurons and mentoring the performance accuracy of the FFNN predictor. Several training functions were examined in this study. Training functions applied through an optimizer. The optimizer iterates over list of input training functions and the ideal algorithm were used in predictor.

Consequently, a three-layered neural network with the structure 22 factors (input layer) × 150 neurons (hidden layer) × 1 (output layer) defined for predictor (Figure 3).

3. RESULT ANALYSIS AND DISCUSSION

The index characteristic values (5829) of valuation samples in the training samples of the study area sent to the FFNN predictor, and their relevant output factors were acquired after network training. These output factors can be utilizing to predict the level of LS in Karabuk. The output FFNN predictor developed based on tabular data from spatial layers in ArcGIS environment. Based on the obtained valuation factors, tabular data (745810 pixels) of whole study area were prepared and migrated to Matlab environment. Then, sent to FFNN predictor and susceptibility values produced as illustrated in Figure 4. Simultaneously, the susceptibility predicting outcomes for landslide were tested and matched over past landslide information. The output demonstrated in Figure 4. Predicted levels of LS in study area can be separated to five categories: very high susceptibility category representing 3.53 km² (3.03 % out of full area); high susceptibility category representing 31.72 km² (27.22% out of full area); moderate susceptibility category representing 27.70 km² (23.77 % out of full area); low susceptibility category representing 22.39 km² (19.22 % out of full area); very low susceptibility category representing 31.16 km² (26.74% out of full area).

In addition, Figure 5 exhibited that the maximum landslide susceptibility category is mostly located within the 423- 1590 m altitude. This category commonly extra prone to landslides than other categories. The outcomes exposed that very high susceptibility category having high correlations with 11 input factors. Correlated factors are: roughness, surface relief ratio, landform, heat load index, dissection, slope/aspect transformation, aspect, general curvature, plan curvature, profile curvature and slope position. The result showed that relevant factors fluctuating in their effects. The maximum LS category is highly influenced zone of the three curvature factors. Spatial analysis examination revealed that close distance zones from rivers streams are highly relevant to the landslides occurrences. Furthermore, heat load index was a significant parameter influencing landslides occurrence.

For operational validation of FFNN predictor for LS, the CM and ROC plots were utilized in this study. CM and ROC plots were used for both training and testing datasets. CM and ROC metrics employed to showed the quality of the prediction FFNN predictor. CM and ROC plots characterize the capability of FFNN predictor to appropriately binary predict LS (occurrence or non-occurrence) from historical defined landslide occasions.

The output from the network during the training stage and actual target were compared. Figure 6 illustrate the performance accuracy of the trained predictor according to confusion matrix. CM showed that the training accuracy is 89.3%, while the accuracy of testing data set is 82.3 accuracy. In addition, Figure 7 present the performance accuracy of FFNN predictor based on Receiver Operating Characteristic (ROC) curves. ROC measure the performance accuracy of the predictor through area under curve. AUC of the training data set is 95.22% and 84.7% for the testing data set. Furthermore, the result showed that there is no overfitting problem in the predictor, since that the performance accuracy in the training data set is higher than the testing data set. The rate of false positive sample is quite small in the training data set and acceptable in testing data. The outcomes of FFNN predictor reflect the successful application of the using NN to perform LS prediction.

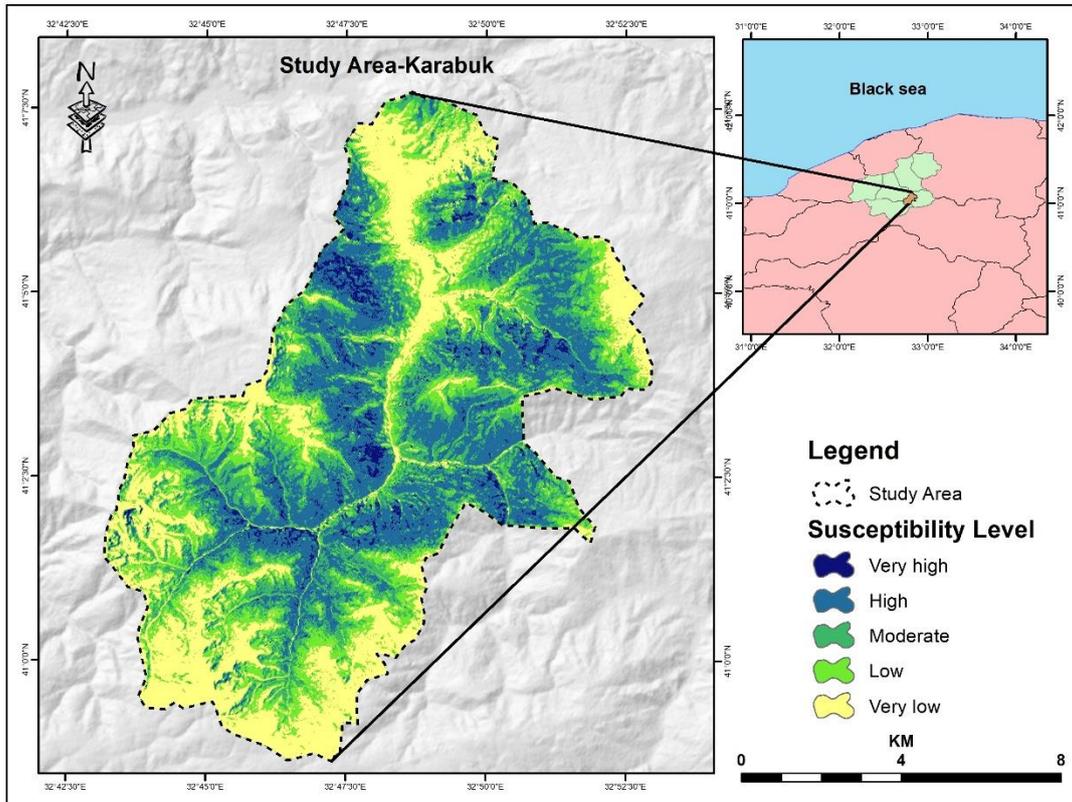


Figure 4 Susceptibility valuation outcomes of landslides based on FFNN predictor.

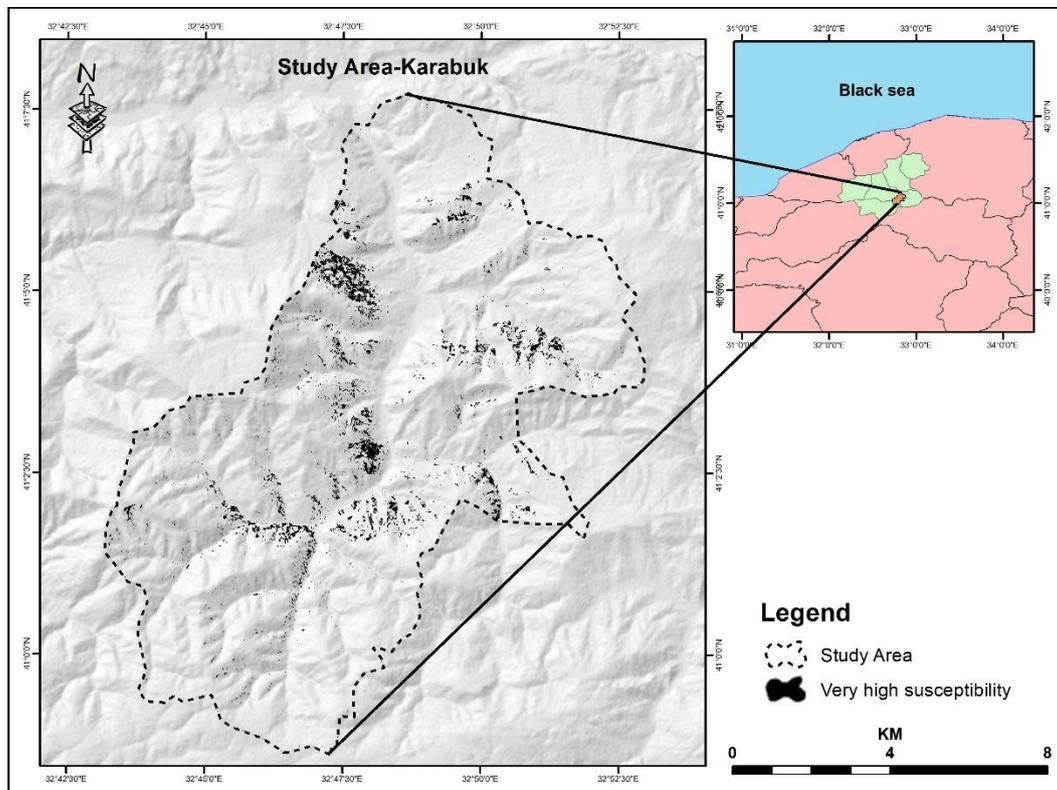


Figure 5 Very high landslides susceptibility category.

Highly performance accuracy of FFNN trained predictor prove the possibility of utilizing only remote sensing satellite images to perform LS prediction. The main advantages of using satellite images in this application area that it is suitable to be implement in difficult topographic, and environmental conditions. For instance, predictor suitable in case of working with mountainous zones, or remote areas that are not accessible or difficult to access. Difficulty could be lack of suitable vehicles or time and financial cost.

Moreover, predictor suitable in case of working with regional or large scale of coverage. One of the main limitation of this point

in others applications frameworks is the difficulty to obtain realistic and suitable data of explanatory parameters. The wide availability and coverage of satellite images are exceedingly support and sustainable this framework. Notably, this framework is suitable to small scale analysis SSA. SSA usually have difficulty to find standardized of consistent spatial resolution for factors. This framework avoided this problem by extracting all factors from the same satellite image. Additionally, in order to use up to date explanatory factors data, it can be easily, low cost and automatically extract it from new satellite images.

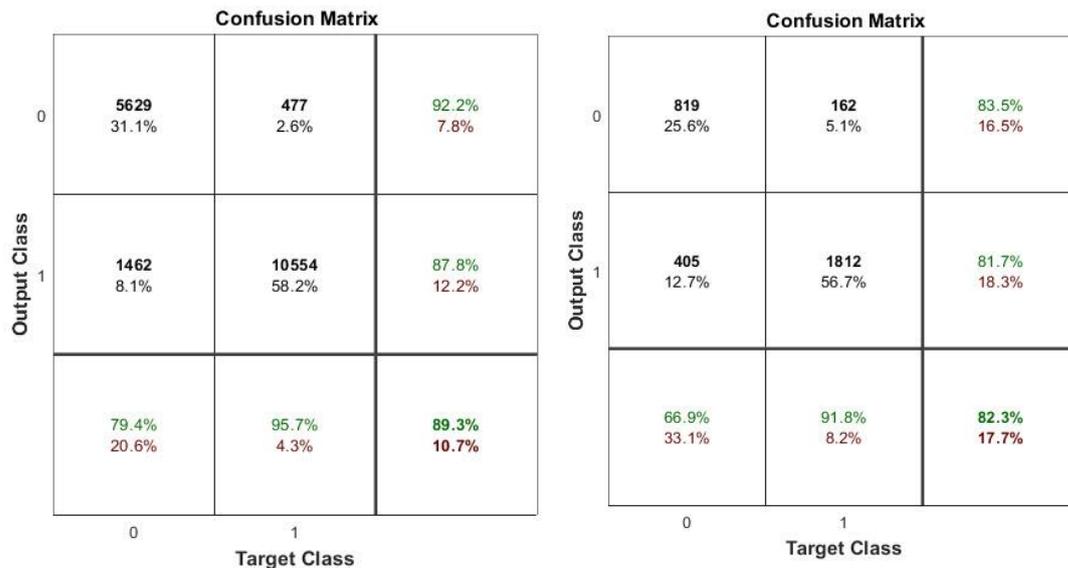


Figure 6 Performance accuracy of FFNN predictor based on confusion matrix.

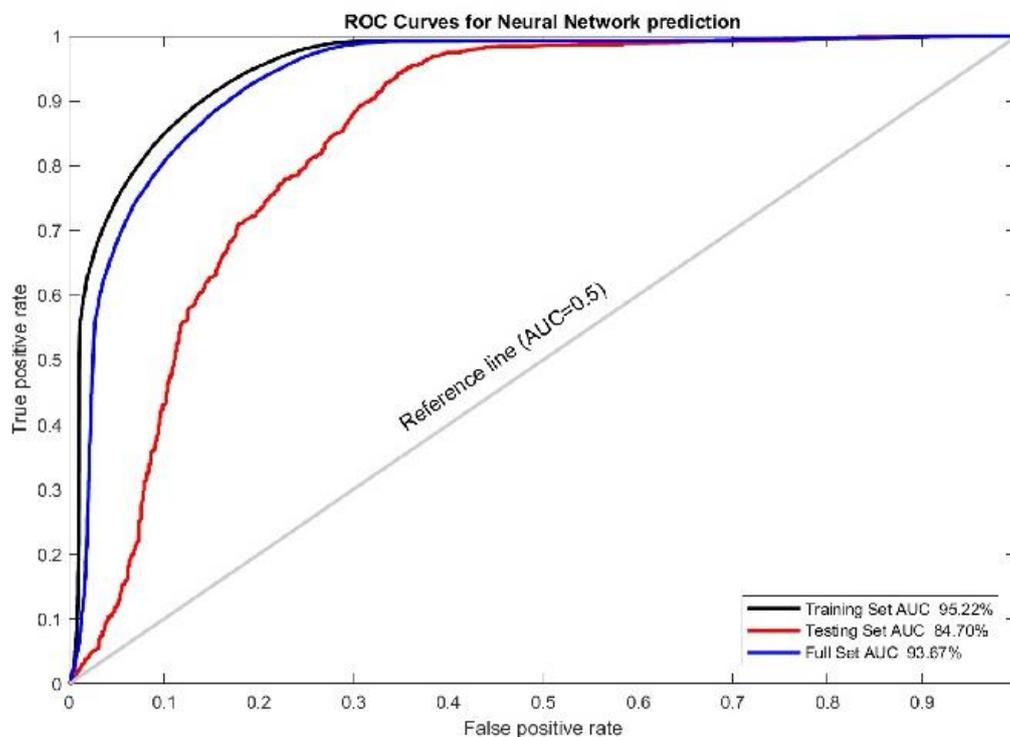


Figure 7 Performance accuracy of FFNN predictor based on Receiver Operating Characteristic (ROC) curves. ROC describe the performance accuracy through area under curve.

On the other hand, this framework consisting of two disadvantages. The first challenge is that FFNN need massive number of neurons in order to gain high performance accuracy, due to the training subjected to the minimum error. In this application, the FFNN predictor utilized 150 neurons in the hidden layer to obtained the good accuracy in Geospatial area. The second challenge is that the FFNN predictor need to extract high number of explanatory factors to increase the index of performance accuracy. These two disadvantages caused high cost of data processing. Especially in case of large-scale analysis, which lead to produce big data. But, since that the application area of LS prediction has medium reproducibility interval (at least weekly). In other words, the analysis repeats every several weeks, not daily or hourly. The disadvantages of the proposed framework can be ignored or handle.

For future works, the authors recommend to LIDAR images for second stage of prediction. Second stage of prediction could only apply based on the very high category susceptibility in this framework.

4. CONCLUSION

Prediction of LS is indispensable in determining prone zones to landslides in Karabuk-Turkey. The main objective of this paper is to propose low data requirement framework for LS prediction. In recent years, several frameworks and methods developed to generate LS maps. The framework consisting three stages; (A) data pre-processing. 3D ALOS PALSAR image utilized.

(B) deriving explanatory factors from the ALOS PALSAR image. From ALOS PALSAR digital elevation model 21 factors were extracted. Factors are: elevation, general curvature, plan curvature, profile curvature, dissection, roughness, landform, slope position, surface area ratio, surface relief ratio, aspect, hillshade, slope, flow direction, flow accumulation, stream order, heat load index, integrated moisture index, site exposure index, compound topographic index, slope/aspect transformation, and 2nd derivative slope. (C) neural networks applied to accomplish LS training and prediction. FFNN neural networks. a three-layered neural network with the structure 22 factors (input layer) \times 150 neurons (hidden layer) \times 1 (output layer). Additionally, 5829 sample points representing past zones of landslide and non-landslide zones were utilized during the training stage. In the last step 745810 points were feed to the trained predictor to create LS map. The generated map of LS showed that 3.53 km² (3.03%) out of the study area located in very high susceptibility category. Performance accuracy were verified by comparing training and testing (89.3% and 82.3 respectively) based on confusion matrix metric. While (95.22% and 84.7% sequentially) based on Receiver Operating Characteristic curve.

To conclude, the framework consequences achieved in this paper exhibited FFNN model can be used as robust predictor for future LS prediction operations. The framework based on low data requirement can predict LS maps. The predicted LS map can be utilized in land administration and departments for future planning and development. The map can be employed to guide the construction and infrastructures projects to the very low susceptibility zones. While instruct them for very high susceptibility for extra geotechnical and geological engineering considerations.

ACKNOWLEDGEMENTS

This study has been supported by 2221 – Fellowship Program of TUBITAK (The Scientific and Technological Research Council of Turkey). We are indebted for their supports.

REFERENCES

- Alaska Satellite Facility's. 2011. "ALOS Dataset Information." Retrieved April 11, 2019 (<https://vertex.daac.asf.alaska.edu/>).
- Astray, Gonzalo, Beatriz Gullón, Jalel Labidi, and Patricia Gullón. 2016. "Comparison between Developed Models Using Response Surface Methodology (RSM) and Artificial Neural Networks (ANNs) with the Purpose to Optimize Oligosaccharide Mixtures Production from Sugar Beet Pulp." *Industrial Crops and Products* 92:290–99.
- Chae, Byung-Gon, Hyuck-Jin Park, Filippo Catani, Alessandro Simoni, and Matteo Berti. 2017. "Landslide Prediction, Monitoring and Early Warning: A Concise Review of State-of-the-Art." *Geosciences Journal* 21(6):1033–70.
- Chaudhary, Pandav, Sachin Kumar Chhetri, Kiran Man Joshi, Basanta Man Shrestha, and Prabin Kayastha. 2015. "Application of an Analytic Hierarchy Process (AHP) in the GIS Interface for Suitable Fire Site Selection: A Case Study from Kathmandu Metropolitan City, Nepal." *Socio-Economic Planning Sciences*.
- Choi, Jaewon, Hyun-Joo Oh, Hong-Jin Lee, Changwook Lee, and Saro Lee. 2012. "Combining Landslide Susceptibility Maps Obtained from Frequency Ratio, Logistic Regression, and Artificial Neural Network Models Using ASTER Images and GIS." *Engineering Geology* 124(0):12–23.
- Conforti, Massimo, Stefania Pascale, Gaetano Robustelli, and Francesco Sdao. 2014. "Evaluation of Prediction Capability of the Artificial Neural Networks for Mapping Landslide Susceptibility in the Turbolo River Catchment (Northern Calabria, Italy)." *CATENA* 113(0):236–50.
- Feliciísimo, Ángel M., Aurora Cuartero, Juan Remondo, and Elia Quirós. 2013. "Mapping Landslide Susceptibility with Logistic Regression, Multiple Adaptive Regression Splines, Classification and Regression Trees, and Maximum Entropy Methods: A Comparative Study." *Landslides* 10(2):175–89.
- Fu, Yingchun, Yaolong Zhao, Yongrui Zhang, Taisheng Guo, Ziwei He, and Jingyi Chen. 2013. "GIS and ANN-Based Spatial Prediction of DOC in River Networks: A Case Study in Dongjiang, Southern China." *Environmental Earth Sciences* 68(5):1495–1505.
- García-Rodríguez, M. J. and J. A. Malpica. 2010. "Assessment of Earthquake-Triggered Landslide Susceptibility in El Salvador Based on an Artificial Neural Network Model." *Natural Hazards and Earth System Science* 10(6):1307–15.
- Japanese Earth observing satellite. 2019. "Advanced Land Observing Satellite - Phased Array Type L-Band Synthetic Aperture Radar." Retrieved April 9, 2019 (https://www.eorc.jaxa.jp/ALOS/en/about/about_index.htm).
- Li, Deren, Shuliang Wang, and Deyi Li. 2015. "Spatial Data Mining."

Liang, Deron, Chih-Fong Tsai, and Hsin-Ting Wu. 2015. "The Effect of Feature Selection on Financial Distress Prediction." *Knowledge-Based Systems* 73:289–97.

Nefeslioglu, H. A., B. T. San, C. Gokceoglu, and T. Y. Duman. 2012. "An Assessment on the Use of Terra ASTER L3A Data in Landslide Susceptibility Mapping." *International Journal of Applied Earth Observation and Geoinformation* 14(1):40–60.

Sener, Erhan, Ozlem Terzi, Sehnaz Sener, and Ramazan Kucukkara. 2012. "Modeling of Water Temperature Based on GIS and ANN Techniques: Case Study of Lake Egirdir (Turkey)." *Ekoloji* 21(83):44–52.

Shahabi, Himan, Saeed Khezri, Baharin Bin Ahmad, and Mazlan Hashim. 2014. "Landslide Susceptibility Mapping at Central Zab Basin, Iran: A Comparison between Analytical Hierarchy Process, Frequency Ratio and Logistic Regression Models." *CATENA* 115(0):55–70.

Song, Kyo-Young, Hyun-Joo Oh, Jaewon Choi, Inhye Park, Changwook Lee, and Saro Lee. 2012. "Prediction of Landslides Using ASTER Imagery and Data Mining Models." *Advances in Space Research* 49(5):978–93.

Song, Shuai, Xilai Zheng, and Fadong Li. 2010. "Surface Water Quality Forecasting Based on ANN and GIS for the Chanzhi Reservoir, China." Pp. 4094–97 in *Information Science and Engineering (ICISE), 2010 2nd International Conference on*. IEEE.

Yeo, In-Ae and Jurng-Jae Yee. 2014. "A Proposal for a Site Location Planning Model of Environmentally Friendly Urban Energy Supply Plants Using an Environment and Energy Geographical Information System (E-GIS) Database (DB) and an Artificial Neural Network (ANN)." *Applied Energy* 119(0):99–117.

Yilmaz, Işık. 2009. "A Case Study from Koyulhisar (Sivas-Turkey) for Landslide Susceptibility Mapping by Artificial Neural Networks." *Bulletin of Engineering Geology and the Environment* 68(3):297–306.

Revised August 2019