

LANDSLIDE SUSCEPTIBILITY MAPPING WITH RANDOM FOREST MODEL FOR ORDU, TURKEY

G. Karakas¹, R. Can¹, S. Kocaman^{1,*}, H. A. Nefeslioglu², C. Gokceoglu²

¹ Dept. of Geomatics Engineering, Hacettepe University, 06800 Beytepe Ankara, Turkey - (gizem.karakas, recepcan, sultankocaman)@hacettepe.edu.tr

² Dept. of Geological Engineering, Hacettepe University, 06800 Beytepe Ankara, Turkey - (hanefeslioglu, cgokce)@hacettepe.edu.tr

ICWG III/IVa

KEY WORDS: Aerial Photogrammetry, Landslide Susceptibility Mapping, Machine Learning, Random Forest, Ordu (Turkey)

ABSTRACT: Landslides are among commonly observed natural hazards all over the world and can be quite destructive for infrastructure and in settlement areas. Their occurrences are often related with extreme meteorological events and seismic activities. Preparation of landslide susceptibility maps is important for disaster mitigation efforts and to increase the resilience. The factors effective on landslide susceptibility map production depend mainly on the topography, land use and the geological characteristics of the region. The up-to-date and accurate data needed for extracting the effective parameters can be obtained by using photogrammetric techniques with high spatial resolution. Data driven ensemble methods are being increasingly used for landslide susceptibility map production and accurate results can be obtained. In this study, regional landslide susceptibility map of a landslide-prone area in a part of Ordu Province in northern Turkey is produced using topographic and lithological parameters by employing the random forest method. An actual landslide inventory delineated manually by geologists using the produced orthophotos and the digital terrain model (DTM) is used for training the model. The results show that an accuracy of 83% and precision of 92% can be obtained from the data and the random forest method. The approach can be applied for generation of regional susceptibility maps semi-automatically.

1. INTRODUCTION

A natural disaster can be defined as event that occurs at unpredictable times, results in loss of life and damages as well as economic losses. Considering the studies on natural hazards in recent years, it can be seen that there is a significant increase in their numbers. Landslide is also one of these natural hazards. For example, a total of 23.041 landslides were enlisted in Turkey (AFAD, 2020). It is necessary to generate landslide inventories and identify landslide prone areas, in order to support regional land use and infrastructure planning and also to increase the awareness for natural hazards and risks. Therefore, preparation of accurate and up-to-date landslide susceptibility maps are of great importance for landslide hazard mitigation efforts.

In August 2018, an extreme meteorological event in Ordu Province of Turkey occurred and caused several deaths and injuries. As a result of this meteorological event, a serious flood occurred. Several structures such as bridges, roads, houses, etc. were collapsed or damaged. During the meteorological event, many landslides (Figure 1) were also triggered that caused further damages on hazelnut gardens, buildings, and people had to be evacuated from their houses due to landslide problem. Ikizce and Caybasi districts of Ordu were also affected heavily and these districts were selected as the study area here.



Figure 1. Example views of the landslides triggered during the extreme meteorological event occurred in Ordu in August 2018. (Image credits: <http://ordu.gov.tr/> (12th August 2018))

* Corresponding author

Aerial photogrammetric techniques can provide the required timely 3D datasets for regional landslide susceptibility and hazard assessments with sufficient accuracy. Unlike the optical satellite images, which are often affected with clouds after a meteorological event, aerial stereo images can be taken after a disaster event. The aerial images are affected less by the clouds and ensure higher flexibility for the time of acquisition. Hence, General Directorate of Mapping (GDM), Turkey, performed an aerial photogrammetric mission right after the flood event and acquired images with a large-format aerial camera, Ultracam Eagle from Vexcel Imaging, Austria, with 30 cm resolution. These photos were used here to produce digital surface model (DSM), digital terrain model (DTM), and orthophotos. The images of the same region were taken previously in 2015 during an ordinary mapping campaign by GDM.

Although these high accuracy datasets are available, the landslides and damages caused by landslides need be assessed manually by human operators. However, manual detection and delineation of landslides require high expertise, can be carried out by engineering geologists and geomorphologists over a long time. An automated or semi-automated process can assist the process to reduce the time and human errors. In addition, modelling of landslides is also important to understand their mechanisms and to model the risks. This task is becoming more and more possible with the availability of high-resolution Earth observation data, both airborne and spaceborne. On the other side, the increase in the data also bring complexity in terms of computing time and power.

Under a collaboration between Hacettepe University, Ankara, and GDM, semi-automated landslide susceptibility and hazard assessment, and modelling procedures have been under development using the 2015 and 2018 Ordu datasets. As a first step, the landslide susceptibility of the region is assessed using a semi-supervised machine learning technique, i.e. random forest and the results are presented here. Nowadays, there has been a significant increase in the implementation of statistical methods and machine learning algorithms for the production of landslide susceptibility maps that are mostly data driven. The major problems with the expert-based methods are the time and labour-intensive processing and sometimes accessibility to the area. The susceptibility assessment results provide the first insight on the areas with landslide potential, which is especially crucial for obtaining rapid results when working with high resolution datasets.

For the aims of the study, the random forest model was selected since it was found very successful in a recent study of the authors (Sevgen et al., 2019) to produce landslide susceptibility map among other machine learning algorithms. Slope gradient, slope orientation, plan and profile curvatures, topographical elevation, lithology, topographic wetness index (TWI) and distance to drainage network were considered as the landslide conditioning factors for the study area. The landslide inventory data were extracted manually using the orthophotos and digital surface model to train the model. The receiver operating characteristics (ROC), including the area under the curve (AUC), was used to assess the accuracy of the model. The result indicates that landslide susceptibility map produced by random forest has good performance (%92) for predicting the future landslides and the susceptibility map can help to identify and analyse the landslide prone areas.

2. METHODOLOGY

2.1 Study Area

The study area includes İkizce and Caybasi Towns of Ordu Province, Turkey. Figure 2 shows the location of the study area and the DTM. A heavy flood occurred in Ordu on August 8th, 2018 caused severe damages to infrastructure and houses. The extents of the flood event were mapped by Tavus et al. (2019, 2020) using Sentinel-1 and Sentinel-2 satellite images of ESA (European Aerospace Agency). Many landslides occurred in the region after the flood.

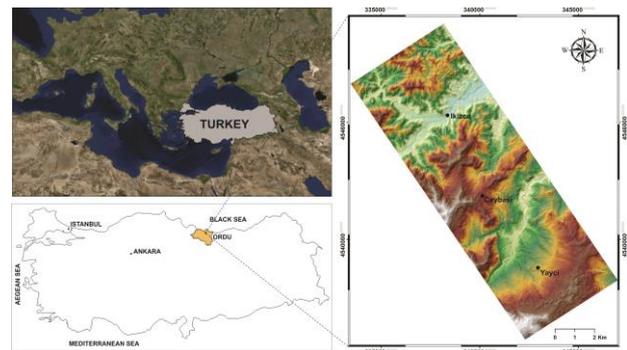


Figure 2. The study area location (left) and the DTM (right) with upper left coordinates: 41,0944°N, 37,0178°E; and the lower right coordinates: 40,9500°N, 37,1814°E.

2.2 Study Workflow

The overall workflow of the study is provided in Figure 3. Aerial photogrammetric flight datasets acquired a few days after the flood by GDM were used to extract the DSM, the DTM and the orthoimages as described in the following section in detail. The geomorphological characteristics of the area were derived from the DTM and used together with the lithology data which was digitized using 1:100,000 scale geological maps published by Altun (2011), as landslide conditioning factors. The actual landslide inventory was prepared manually by using the high resolution DTM and the orthophotos. The random forest method (Breiman, 2001) was employed as the prediction method for landslide map production by using the landslide inventory for training. The output map was validated with respect to the test samples (validation data), which were not included in the model training step. Details of the data pre-processing and the random forest method are provided in the following sections.

2.3 Preparation of Photogrammetric Datasets

A total of 11 aerial stereo images with 30 cm ground sampling distance (GSD) were used to produce the DSM, the DTM and the orthophotos. The images were acquired by GDM using Ultracam Eagle camera in a flight mission in 2018 after the flood occurrence, which triggered several landslides. The image set used in this study form a single flight path (strip). The Ultracam Eagle camera used during the mission has an image frame format of 20010 x 13080 pixels with 0.005 mm detector pixel size and 100.5 mm focal distance. The interior orientation parameters (i.e. camera calibration data) and the exterior orientation parameters estimated in a bundle block

adjustment process using GNSS surveyed ground control points were obtained from the GDM. Trimble Inpho software (Trimble, 2020) was employed for the generation of orthophotos with 30 cm GSD, and the DSM and the DTM with 1m grid intervals. It must be noted that the Inpho software provides the options for producing both the DSM and the DTM, which applies a filter for the latter one. The expected accuracy of the point positioning from the data is ca. 15 cm in planimetry and in height.

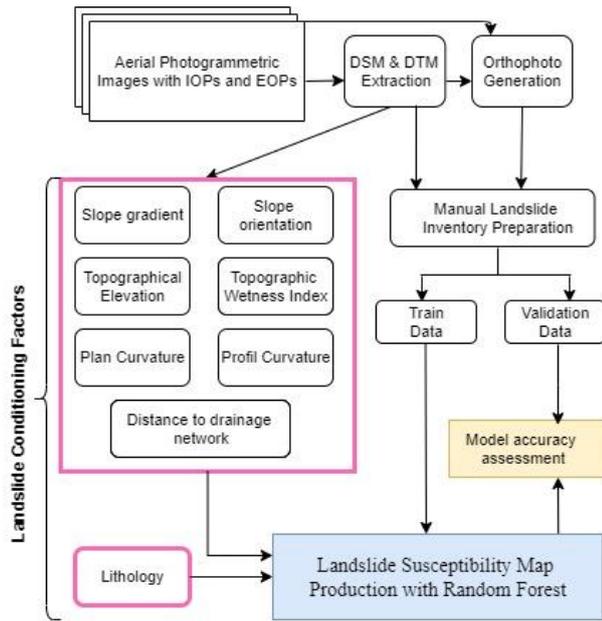


Figure 3. The study workflow.

2.4 Geological Characteristics of the Site and the Landslide Inventory

The stratigraphic sequence begins with the Campanian-aged trachyandesite, andesite and rhyodacites and continues with Maastrichtian-Palaeocene aged mudstone, limestone, sandstone and marl in the study area (Figure 4). Maastrichtian aged limestones overlie Maastrichtian-Palaeocene units in the region. Volcanic rocks developed due to the volcanism active during the Middle Eocene and Late Eocene periods are also observed in the study area. The Early-Middle Eocene aged sandstone, mudstone and limestone units, which deposited simultaneously with this volcanic activity, are observed together with the Middle-Late Eocene aged andesite, basalt and pyroclastic rocks (Altun, 2011).

The landslides observed in the study area were mapped using the DSM and orthophotos produced in this study. Total 25 landslides were mapped manually (Figure 5). The movements were classified as deep-seated circular active failures by considering the characteristics suggested by Cruden and Varnes (1996). The minimum and maximum landslide area values were calculated as 0.4 km² and 5.4 km², respectively. The total landslide area was calculated as 17 km².

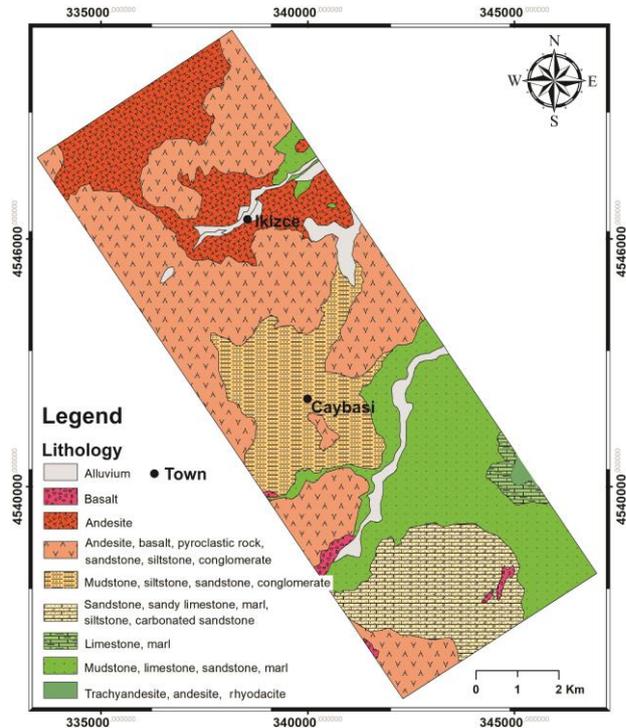


Figure 4. Lithological map of the study area (Altun, 2011).

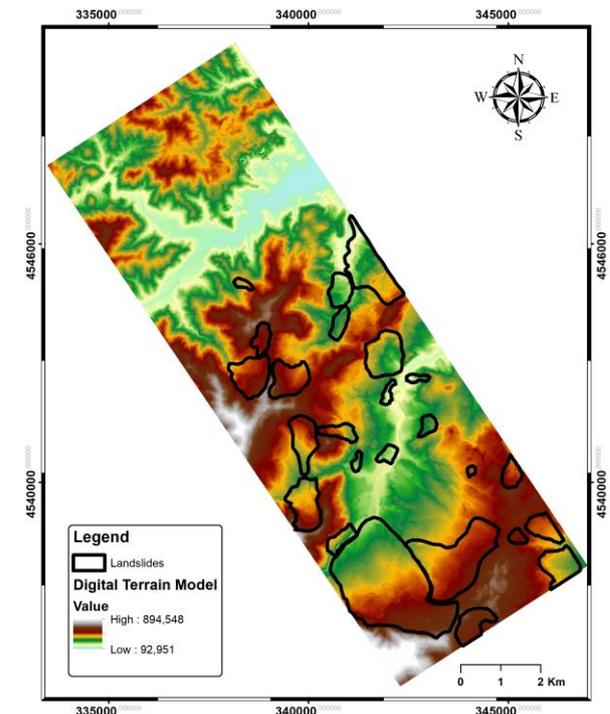


Figure 5. Landslide inventory and the DTM.

2.5 Landslide Conditioning Factors

To evaluate landslide susceptibility in the region the conditioning factors were investigated. For this purpose, the lithological distribution and topographic factors with respect to landslides were assessed. The most prone lithology observed in the region is Early-Middle Eocene aged sandstone, mudstone, and limestones. Approximately 40% of the failures were mapped in this unit. Additionally, 20% of the failures were observed on Early-Middle Eocene aged sandstone and mudstone, 20% of the failures were observed on Maastrichtian-Palaeocene aged mudstone, limestone, sandstone and marl, and 16% of the failures observed on Middle-Late Eocene aged andesite, basalt and pyroclastic rocks. Totally six different topographic parameters, altitude as DTM, slope gradient, slope aspect, slope curvatures plan and profile, topographic wetness index and a hydrological factor distance to drainage network were investigated as the landslide conditioning topographic parameters in the region (Figures 6-11). A statistical summary of these parameters for the whole study area and the area covered by the landslide inventory polygons are provided in Tables 1 and 2, respectively. Accordingly, landslides within the study area are observed on the mean topographic slopes of 17 degrees, in areas with a mean TWI value of 4.55, and at a mean distance of approximately 406 m from the drainage network.

Table 1. Statistical summary of the topographical data obtained for the whole study dataset.

	Min	Max	Mean	Std. Dev.
DTM (meters)	92.95	894.54	386.63	130.89
Slope (degrees)	0	86.85	23.11	13.89
Aspect (degrees)	-1	360	179.03	106.86
TWI	-3.16	25.31	4.29	2.45
Distance to Drainage Network (meters)	0	1706.05	477.82	342.34
Plan Curvature	-3.73	6.35	3.50	7.91
Profil Curvature	-4.41	5.17	-8.81	6.14

Table 2. Statistical summary of the topographical data obtained for the landslide inventory area dataset.

	Min	Max	Mean	Std. Dev.
DTM (meters)	151.77	703.55	412.91	96.04
Slope (degrees)	0	77.11	17.59	12.32
Aspect (degrees)	-1	360	194.42	112.85
TWI	-1.70	25.23	4.55	2.65
Distance to Drainage Network (meters)	0	1661.79	406.68	300.70
Plan Curvature	-1.22	1.44	3.58	6.60
Profil Curvature	-2.30	1.93	-1.54	6.10

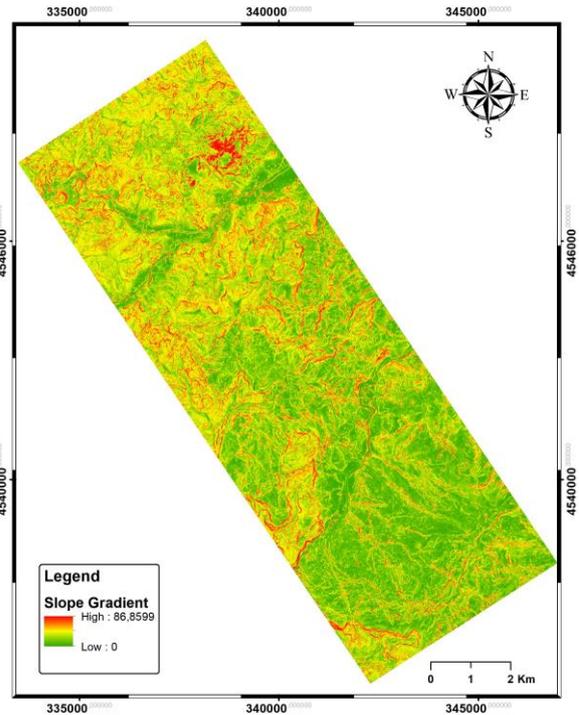


Figure 6. Slope gradient map of the study area.

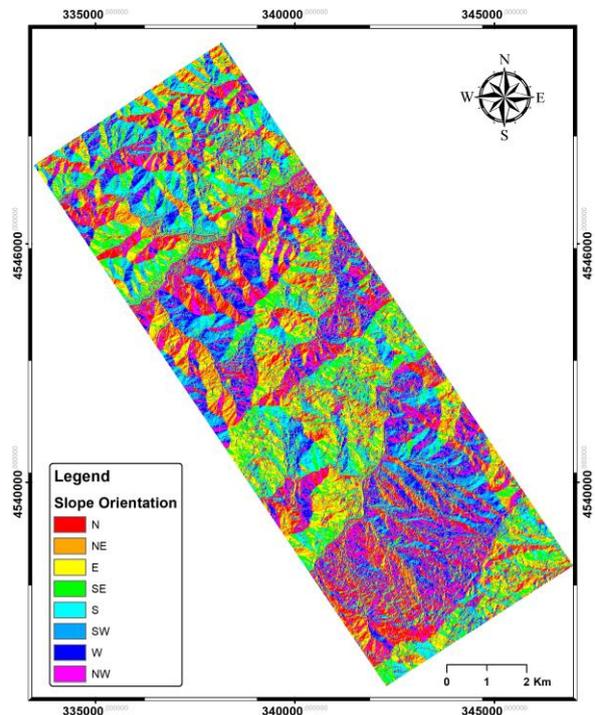


Figure 7. Slope aspect map of the study area.

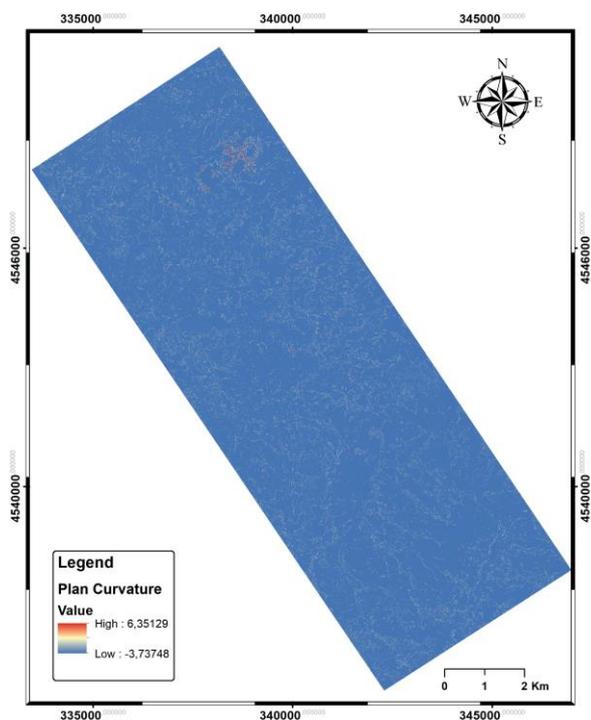


Figure 8. Plan slope curvature map of the study area.

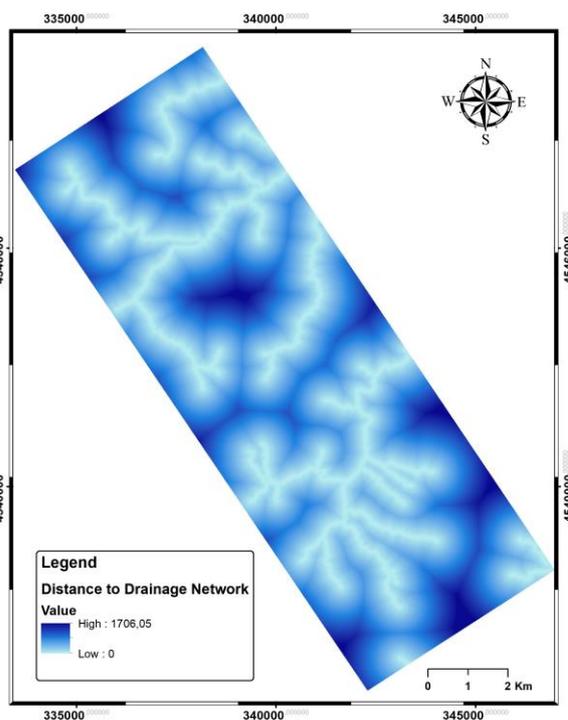


Figure 10. Distance to drainage network map of the study area.

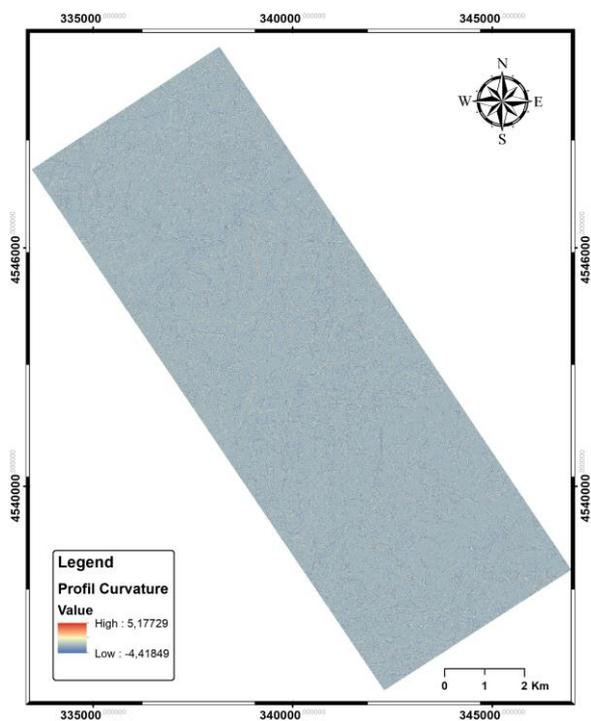


Figure 9. Profile slope curvature map of the study area.

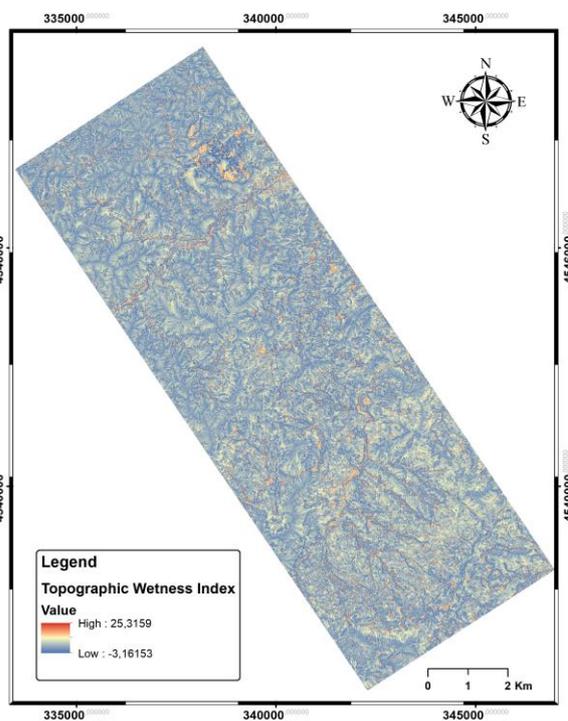


Figure 11. Distance to drainage network map of the study area.

2.6 Landslide Susceptibility Mapping with the Random Forest Method

The random forest is an ensemble method of decision trees (DT) (Breiman, 2001; Sevgen et al., 2019). The DT is created randomly with the method at the training stage; and the DTs are evaluated for the best score based on the average of the results of the trees. The random forest tries to select the most important features when creating the DTs. The use of the random forest method is relatively new for landslide susceptibility mapping (e.g. Hong et al., 2016; Dou et al., 2019; Chen et al., 2017; Chu et al., 2019; Sevgen et al., 2019).

In the present study, using the geological and topographic factors, the landslide susceptibility map of the study area was produced with the random forest method. Due to the uneven distribution of the landslide inventory in the study area, only the area in the middle part of the site (marked with blue rectangle in Figure 12) was used for the model training using the landslide and non-slide samples. The landslide samples were selected from the red polygons in Figure 12. The larger landslides in the southern part of the area and the other landslides in the East were not employed in the model training stage (depicted with green polygons in Figure 12). However, they were utilized for the accuracy assessment. The model training results were applied to the whole dataset to produce the output landslide susceptibility map as shown in Figure 13. The classes were formed with probabilities of equal interval.

Python scikit-learn (Scikit-learn, 2020), a free and open source library, is used for performing the random forest method. The `sklearn.ensemble.RandomForestClassifier` function of the library is used with the following input parameters:

```
n_estimators=128,
criterion='entropy',
max_depth=16,
min_samples_split=2,
min_samples_leaf=4,
class_weight='balanced',
bootstrap='true',
random_state=32,
oob_score='true',
n_jobs=-1
```

Although the input values were selected heuristically to obtain the preliminary results, Grid Search Cross Validation method is planned to be investigated in the future. Considering the resolution of the datasets, computation optimization is also necessary. However, the level of accuracy and precision as described below has also been found satisfactory. A total of 139.096.240 pixels (8 feature classes, each composed of 6.954.812 landslide and 10.432.218 non-landslide pixels) was used in the training stage. The total number of pixels in the study area is 88.790.485.

The performance assessment of the model was also investigated with the ROC (receiver operating characteristics) curve and the AUC (area under curve) statistics (Swets, 1998) and the accuracy. For the model training, the landslide polygons in Figure 12a were used with 80/20 ratio and test pixels were used for the ROC graph. The area under curve value was obtained to be 0.92 indicating that the model is successful enough to predict possible future landslide

occurrences in the region (Figure 14). In the ROC curve, classes 0 and 1 indicate the non-landslide and landslide data, respectively. The accuracy assessment result obtained from the test dataset was 83%.

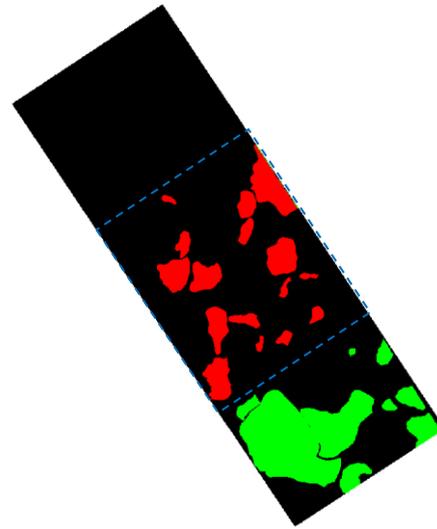


Figure 12. The model training area (blue rectangle) with the landslide inventory used for the training (red polygons). The landslides depicted with green polygons were used for the accuracy assessment.

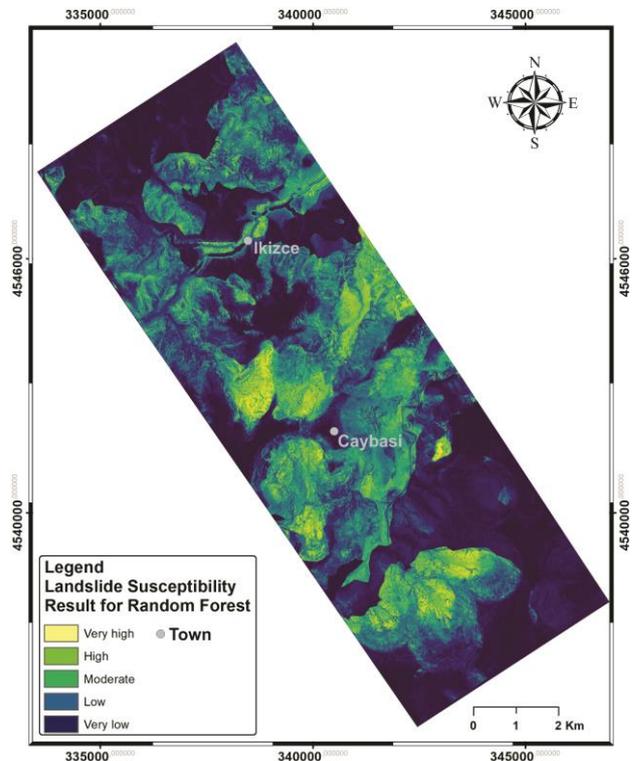


Figure 13. Landslide susceptibility map of the study area.

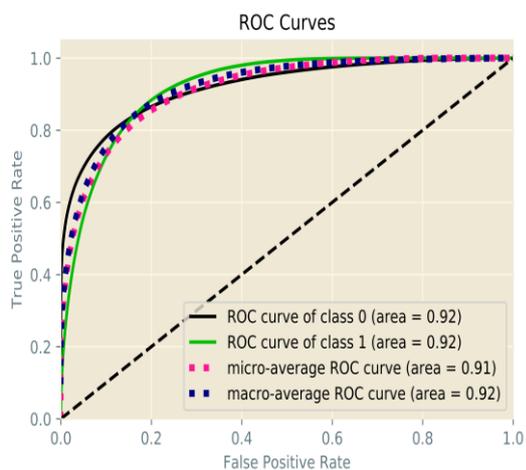


Figure 14. The Receiver Operating Characteristics (ROC) curve evaluation of the model.

3. DISCUSSIONS AND CONCLUSIONS

Ordu Province in general is one of the most landslide prone area in Turkey. One of the fundamental precautions to be taken against landslides is the mapping of landslide hazard. In other words, calculating spatial probabilities related to landslides likely to develop in the future is one of the most basic stages. This process, also known as landslide susceptibility mapping, requires up-to-date data. In this sense, current orthophoto production provides an advantage in this regard. Furthermore, the use of models with high predictive capacity in the development of landslide susceptibility models increases the reliability of the landslide susceptibility maps produced.

In this study, in a highly landslide prone area, current orthophoto, DSM and DTM productions were made and a landslide susceptibility model was developed using the random forest method. The landslide susceptibility map produced using the developed model was evaluated in 5 equally divided classes as very low, low, moderate, high and very high (Table 3). The area with high and very high susceptibility values within the study area is obtained as 10.89 km² and 3.44 km² respectively.

Table 3. The landslide probability distribution in the study area.

Class	Probability range (%)	Area (km ²)	Percentage
Very High	80-100	3.44	3.87%
High	60-80	10.89	12.27%
Moderate	40-60	15.24	17.16%
Low	20-40	17.13	19.29%
Very low	0-20	42.09	47.41%

The landslide susceptibility map produced here will be used to determine possible landslide areas in different parts of Ordu to develop rule sets in rule-based classifications for the future studies. In addition, a combined analysis with the results of this study and the flood extent map produced by Tavus et al. (2019, 2020) will be performed to investigate the relationship between the flood and the landslide events. Although a method

for multi-hazard susceptibility assessment for flood and landslide has been proposed by Yanar et al. (2020) previously, more future studies are needed in this field to comprehend the nature of the natural hazards.

ACKNOWLEDGEMENTS

The authors gratefully thank to General Directorate of Mapping, Turkey, for provision of aerial photogrammetric datasets and Dr. Orhan Firat for his tireless efforts and support.

REFERENCES

- AFAD. Afet ve Acil Durum Yönetimi Başkanlığı. URL: <https://www.afad.gov.tr/> (accessed on 30 April 2020).
- Altun, I.E., 2011. Maden Tetkik ve Arama Genel Müdürlüğü 1:100000 ölçekli Türkiye Jeoloji Haritaları F38 ve G38 Paftaları, No:151, Ankara, Türkiye.
- Breiman, L., 2001. Random Forests. *Machine Learning*, 45, 5–32.
- Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Tien Bui, D., Duan, Z., Ma, J., 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, 151, 147–160.
- Chu, L., Wang, L.J., Jiang, J., Liu, X., Sawada, K., Zhang, J., 2019. Comparison of landslide susceptibility maps using random forest and multivariate adaptive regression spline models in combination with catchment map units. *Geosciences Journal*, 23, 341–355.
- Cruden, D.M., Varnes, D.J., 1996. Landslide types and processes. In: Turner, A.K., Schuster, R.L. (Eds.), *Landslides: Investigation and Mitigation*, pp. 36–75 Transp Res. Board, Nat Res. Council Spec Rep 247.
- Dou, J., Yunus, A.P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.W., Khosravi, K., Yang, Y., Thai Pham, B., 2019. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Science of the Total Environment*, 662, 332–346.
- Hong, H.Y., Pourghasemi, H.R., Pourtaghi, Z.S., 2016. Landslide susceptibility assessment in Lianhua County (China): A comparison between random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* 2016, 259, 105–118.
- Scikit-learn, 2020. Python Library. <https://scikit-learn.org/stable/> (last accessed on 03.05.2020)
- Sevgen, E., Kocaman, S., Nefeslioglu, H.A., Gokceoglu, C., 2019. A Novel Performance Assessment Approach Using Photogrammetric Techniques for Landslide Susceptibility Mapping with Logistic Regression, ANN and Random Forest. *Sensors*, 19, 3940.

Swets, J.A., 1998. Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293.

Tavus, B., Kocaman, S., Nefeslioglu, H.A., Gokceoglu, C., 2019. Flood Mapping Using Sentinel-1 SAR Data: A Case Study of Ordu 8 August 2018 Flood. Bi-annual Symposium of Turkish Society for Photogrammetry and Remote Sensing, April 25-27, Aksaray, Turkey.

Tavus, B., Kocaman, S., Gokceoglu, C., Nefeslioglu, H.A., 2020. A Fusion Approach for Flood Mapping Using Sentinel-1 and Sentinel-2 Datasets. ISPRS Congress 2020 (accepted for publication).

Trimble, 2020. Inpho Photogrammetry Suite. <https://geospatial.trimble.com/products-and-solutions/inpho> (accessed on 30 April 2020)

Yanar, T., Kocaman, S., Gokceoglu, C., 2019. On the Use of Sentinel-2 Images and High Resolution DTM for Landslide Susceptibility Mapping in a Developing Urban Settlement (Mamak, Ankara, Turkey). *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-3/W8, p469-476. <https://doi.org/10.5194/isprs-archives-XLII-3-W8-469-2019>

Yanar, T., Kocaman, S., Gokceoglu, C., 2020. Use of Mamdani Fuzzy Algorithm for Multi-Hazard Susceptibility Assessment in a Developing Urban Settlement (Mamak, Ankara, Turkey). *ISPRS Int. J. Geo-Inf.*, 9, 114.