Forest fire susceptibility assessment with machine learning methods in north-east Turkiye

Conference Paper

Author(s):
Kantarcioglu, Omer; Schindler, Konrad; Kocaman, Sultan (iD)

FOREST FIRE SUSCEPTIBILITY ASSESSMENT WITH MACHINE LEARNING METHODS IN NORTH-EAST TURKIYE

O. Kantarcioglu [1,2], K. Schindler [3], S. Kocaman [2*]

[1] Hacettepe University, Graduate School of Science and Engineering, Ankara, Türkiye – omerkantarcioglu@hacettepe.edu.tr
[2] Dept. of Geomatics Engineering, Hacettepe University, 06800 Beytepe Ankara, Türkiye - sultankocaman@hacettepe.edu.tr
[3] ETH Zurich, Photogrammetry and Remote Sensing, 8092 Zurich, Switzerland – schindler@ethz.ch

KEY WORDS: Forest fire susceptibility, forest inventory, random forest, artificial neural network, spatial probability distribution

ABSTRACT:

Forest fires have devastating effects on biodiversity, climate, and humans. Producing detailed and reliable forest fire susceptibility maps is crucial for disaster management. Data-driven machine learning methods can be applied for forest fire susceptibility mapping, and learning data required for this purpose can be obtained from high-resolution satellite imagery along with a fire inventory. In this study, we assessed the performances of Random Forest (RF) and artificial neural network (ANN) classifiers for producing forest fire susceptibility maps of a region in north-east Türkiye covering Trabzon, Gümüşhane, Rize, and Bayburt provinces using freely available Earth observation data and forest inventory provided by the regional directorate. Forest type, EU-DEM v1.1 (25 m), and tree cover density were retrieved from Copernicus Land Monitoring Service. Sentinel-2 images were utilized for calculating spectral indices such as normalized difference vegetation index and modified normalized difference water index to assess surface water and vegetation characteristics. Thus, a total of twelve variables including topographic, anthropogenic, hydrologic, vegetation and land use data were used as input. The RF and ANN illustrated similar prediction performances based on receiver operating characteristics (ROC) area under the curve (AUC) values, which were 0.89 and 0.88, respectively. The RF performed better in terms of overall accuracy and F-1 score. The susceptibility maps with 25 m resolution were also investigated visually. The ANN results predicted higher susceptibility levels and larger areas were found prone to wildfire. Leave-one-out analysis results indicated that elevation was the most influential factor based on the achieved OA.

1. INTRODUCTION

Forest fires have devastating effects on a variety of environmental, social, and economic resources (Gill et al. 2013) Additionally, forest fires cause air pollution and diminish its quality, destroy forest ecosystems, and lead to global warming (Leuenberger et al. 2018). Therefore, the development of an efficient disaster management system to decrease the severity of wildfires and support mitigation efforts is essential (Adab et al. 2013). Although a variety of methods has been utilized in the literature to assess forest fire susceptibility, and the number of studies has increased, further investigations are still needed since the model accuracy depends on the method applied and data availability/quality. More than 30 methods were used for this purpose between 2001 and 2021 (Chicas and Østergaard Nielsen 2022). The most frequently used data-driven machine learning (ML) models have been the random forest (RF) model (Cao et al. 2017) proposed by Breiman (2001), logistic regression (LR) (Rodrigues and de la Riva 2014), support vector machines (SVM) (Sachdeva et al. 2018), and artificial neural networks (ANN) (Zhang et al. 2021; Kantarcioglu et al., 2023). Comprehensive reviews on wildfire susceptibility and hazard mapping can be found in the literature (Jain et al., 2020; Chicas & Østergaard Nielsen, 2022).

The number and severity of forest fires have increased drastically in Turkiye over the last few years. Akinci and Akinci (2023) assessed the forest fire susceptibility for Manavgat region of Antalya province, Turkiye using the ignition inventory of local forest authority and the XGBoost algorithm, based on boosting decision trees. Another recent work by Kantarcioglu et al. (2023) employed an ANN to produce susceptibility map of Thrace region, Turkiye and compared with the RF. An extensive forest database with tree types, tree ages, forest roads, and ignition locations of 1506 fires over a period of eight years was utilized in the study, and the forest roads was found the most predictive variable.

In this study, two widely used ML classifiers, RF and ANN, were again implemented for producing forest fire susceptibility maps for the north-east side of Türkiye covering Trabzon, Gümüşhane, Rize, and Bayburt provinces. The ignition locations were obtained from Trabzon Regional Directorate of Forestry (TRFD). The inventory includes 368 fire events recorded between the years 2013 and 2022. Thus, this study aims to identify areas prone to forest fires (ignition) by predicting their spatial probability using various input variables in combination with RF and ANN. Based on the literature survey, analysis of regional characteristics, and data availability, a total of 12 features were identified, including elevation, aspect, slope, tree cover density, forest type, settlement proximity, settlement density, water proximity, power line proximity, normalized vegetation difference index (NDVI), modified normalized water difference index (MNDWI), and land use land cover (LULC). The susceptibility maps obtained from both models were inspected visually and validated by using receiver operating characteristics (ROC) area under the curve (AUC) values, overall accuracy (OA), and F-1 score. In addition, feature importance analysis based on leave-one-out approach is presented and discussed.

---

* Corresponding author

2. MATERIALS AND METHODS

2.1 Study Area and Fire Characteristics

The study site is located in the eastern part of Black Sea region, Turkiye, and covers an area of 18,877 km^2, Figure 1. The area is rather mountainous, and the northern parts are located at higher altitudes as can be seen from the EU-DEM v1.1 (2023) data, Figure 2. Trabzon and Rize are more vegetated compared to Gümüşhane and Bayburt, according to European Space Agency (ESA) WorldCover (2023) data, Figure 3. The Pontic mountains located in the north of the region differentiate the climate and vegetation characteristics of the provinces.
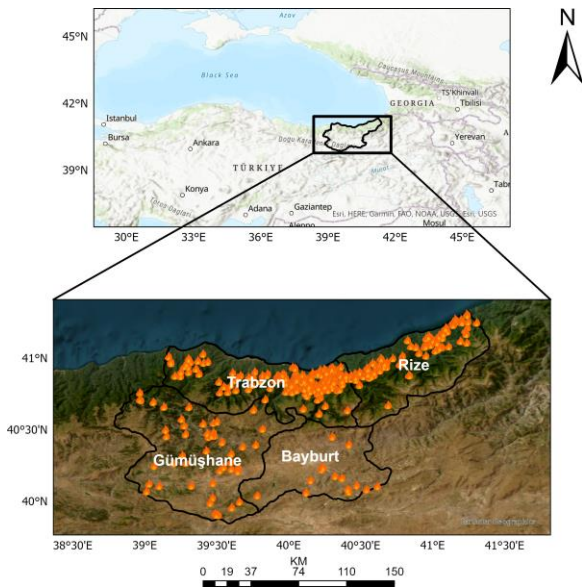

Figure 1. Location map of the study area, and ground truth fire ignition points.
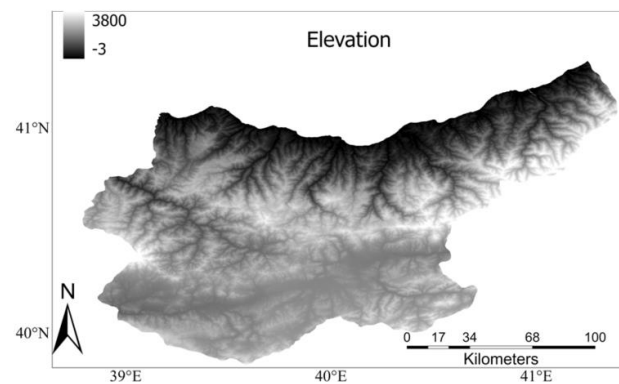

Figure 2. Altitudes within the study area, from EU-DEM v1.1 (2023).

The fire inventory used here consists of locations and dates of ignition events that occurred between 2013 and 2022 (see Figure 1). Further attributes involve risk level description conducted by the emergency response teams and the size of the burned areas, which can be used for hazard assessments in the future. A total of 368 fires served as the basis for forest fire susceptibility mapping. Half of these fires resulted in burnt areas < 1 hectare. The largest burnt area is approximately 130 hectares.
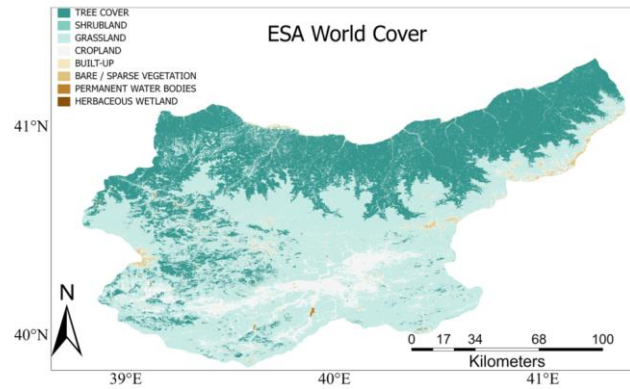

Figure 3. LULC map of the study area, from ESA WorldCover data (ESA WorldCover, 2023).

2.2 Methodology

The overall workflow of the study is presented in Figure 4. It consists of the preparation of the input features, selection of training and test samples, application and validation of the ML models (RF or ANN), and the production of susceptibility maps by applying the learned models. Further details are provided below.
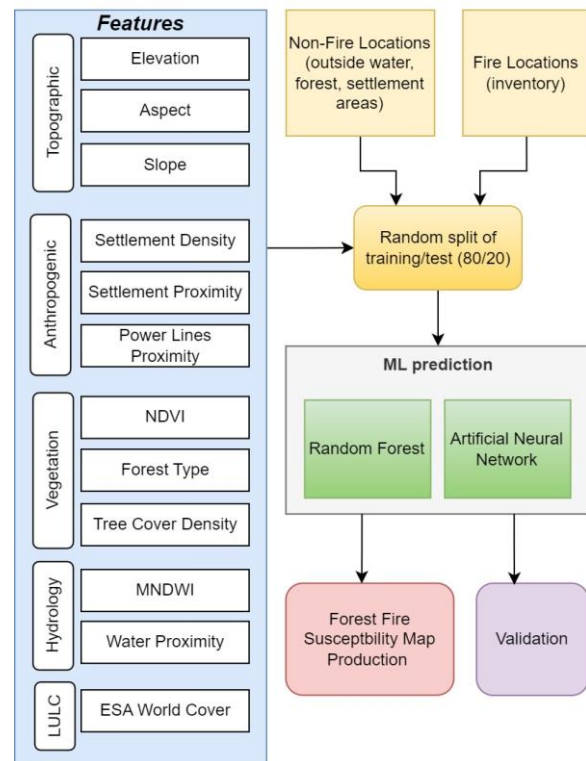

Figure 4. The workflow of the study.

2.2.1 Input Datasets and Pre-processing: A total of six images obtained from Sentinel-2 satellites of ESA were mosaicked and clipped for the site to produce maps of NDVI and MNDWI. All Sentinel-2 images were acquired on the 29th September 2020. As digital elevation model (DEM), EU-DEM v1.1 (2023) with 25 m resolution was employed, additionally slope and the aspect maps were produced using this dataset. Tree cover density and forest type data were obtained from the Copernicus Land Monitoring Service (CLMS, 2023). In

addition, power line data were extracted from World Bank global electricity transmission and distribution lines (World Bank, 2023). Buildings in the study area were extracted from OpenStreetMap (OSM) Project repository (OSM, 2023). Likewise, water surfaces were extracted from the EU-Hydro database provided by EU Copernicus Programme (EU-Hydro, 2023). A brief summary of the datasets is given in Table 1.

Input Data	Source	Format	SR	Date
DEM	EU EEA[1]	Raster	25 m	2011
Forest Type	EU EEA[1]	Raster	10 m	2018
Tree Cover Density	EU EEA[1]	Raster	10 m	2018
Settlements	OSM	Vector	-	2022
LULC	ESA World Cover[1]	Raster	10 m	2021
Power Lines	Worldbank[2]	Geopackage	-	
Water	EU Hydro[1]			
Sentinel-2 imagery	ESA[1]	Raster	10 m	2020
Fire Inventory	TRDF[3]	Vector	-	2013-2022
Buildings	OSM	Vector	-	2022

[1] *via EU Copernicus Service (CLMS, 2023);* [2] *(Arderne et al., 2020);* [3] *Trabzon Regional Directorate of Forestry (2022)*

Table 1. Input data characteristics.

The data pre-processing was carried out using ArcGIS Pro software by Environmental Systems Research Institute (ESRI). The datasets were clipped according to the study area boundaries. The vector datasets (settlements, buildings, power lines, river network) were rasterized w.r.t. the grid interval of EU-DEM. Density maps were estimated using the kernel density method. Proximity features were derived according to the Euclidean distance. MNDWI and NDVI features were computed from the Sentinel-2 imagery to assess the vegetation and water areas in the study area using Equations 1 and 2.

$$NDVI = (NIR - Red) / (NIR + Red), \qquad (1)$$

$$MNDWI = (Green - SWIR) / (Green + SWIR). \qquad (2)$$

The negative (non-fire) reference samples were randomly selected. Their number was set 50% greater than fire samples for a ratio of 1:1.5, to avoid any serious class imbalance. The samples were randomly split as training (80%) and validation (20%) sets (Figure 5). Existing Python libraries were utilized for the implementation of the RF and ANN, with manual hyperparameter optimization as explained next.

2.2.2 ANN: An ANN consists of layers of neurons that each implement a linear combination followed by a non-linear activation function (Negnevitsky, 2002). The hyperparameters of the ANN model trained here were set manually, based on the susceptibility map and the AUC value of the result. Keras and Tensorflow (Abadi et al. 2016) libraries were used to implement the ANN method. In ANN, two hidden layers were used. These hidden layers were activated with the "relu" function. The output of these hidden layers was activated with the "sigmoid" function to calculate the predictions. Lastly, for compiling the ANN, the "adam" optimizer parameter was used.
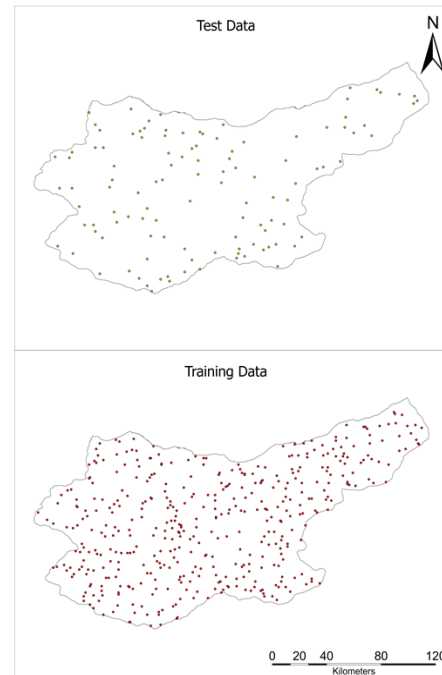

Figure 5. The distribution of training and test samples.

2.2.3 RF: The RF (Breiman 2001) is a supervised classification method consisting of an ensemble of decision trees. Each tree classifies the input according to a sequence of thresholds on the input features, learned from the training data. The RF was implemented using Sci-kit Python libraries (Pedregosa et al. 2011). Parameter selection has been performed via manual testing based on the accuracy results. In RF, the maximum depth of the tree is set as sixteen, the minimum number of samples required to be at a leaf node is four and the minimum number of samples at the leaves is twelve.

2.2.4 Validation: The test data was used for performance assessment via ROC curve, which contrasts recall or sensitivity (Eq. 3) and specificity (Eq. 4). F-1 score (Eq. 5) and OA value (Eq. 6) were also computed. Visual inspection of the predicted maps was carried out by the authors as a further assessment. The importance of different features was analyzed using leave-one-out approach, which is based on removal of one feature from the full set at the learning stage. As a further analysis, the importance of the features was also calculated using

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (4)$$

$$F1 = \frac{2xTP}{2xTP + FP + FN} \qquad (5)$$

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \qquad (6)$$

where TP is true positive, FN is false negative, TN is true negative, and FP is false positive.

3. RESULTS AND DISCUSSIONS

In the following, the input feature maps obtained from the pre-processing, the susceptibility maps produced with the RF and the ANN, the validation results, and the feature importance analysis are presented and discussed.

3.1 Input Features: The input features obtained from data pre-processing were categorized as topographic, anthropogenic, LULC, vegetation and hydrological (Table 2). Feature maps are shown in Figure A1 in the Appendix.

Topographic	Anthropogenic	LULC	Vegetation	Hydrology
Elevation	Settlement Density	ESA Word Cover	NDVI	Water Proximity
Aspect	Settlement Proximity		Forest Type	MNDWI
Slope	Power Lines Proximity		Tree Cover Density	

Table 2. Input features implemented in the study.

3.2 Forest Fire Susceptibility Maps: The maps obtained from both the ANN and the RF methods are shown in Figure 6. The RF predicted higher susceptibility values (up to 0.96). Both maps show similar spatial distributions overall. However, there are local differences in different land cover types (water, forest, roads, settlements), as illustrated in Figure 7. The map details shown in the figure also indicate that the ANN predicts larger areas as fire-prone.
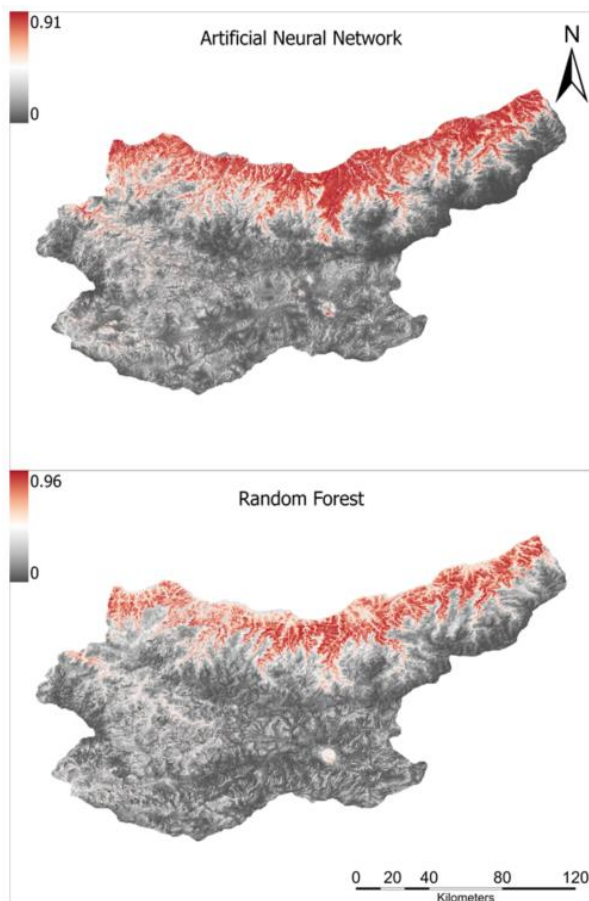

Figure 6. Forest fire susceptibility maps obtained from the ANN (above) and the RF (below).
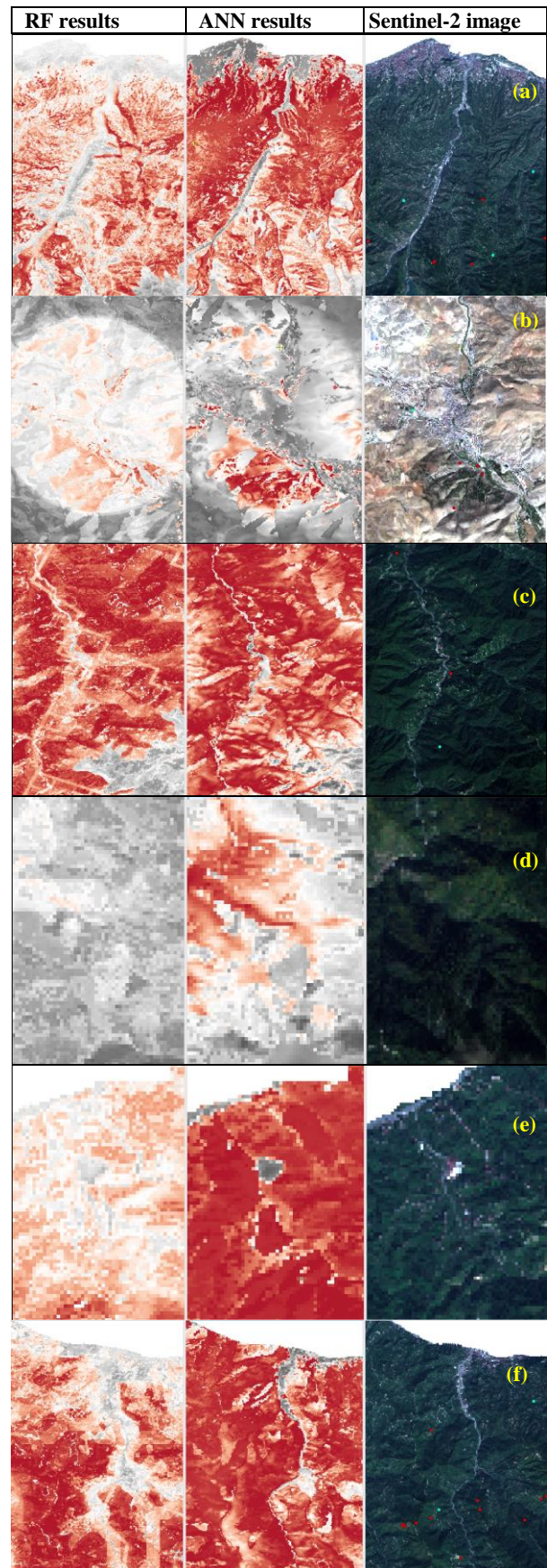

Figure 7. Local details of the maps shown in Figure 6. (a) Trabzon city (b) Bayburt city; (c) Hemşin, Rize; (d) Kızılot, Gümüşhane; (e) Isıklı Village, Ardeşen; (f) Sürmene, Trabzon.

3.3 Model Validation Results: The AUC values obtained from RF and ANN were 0.89 and 0.88, respectively, Figure 8. The corresponding OA were 0.86 and 0.80, and the F1 scores were 0.81 and 0.75. The RF exhibited better prediction performance based on F-1 score and the OA, which indicates a better fit of the model to the learning data. On the other hand, the ANN found larger areas as fire-prone (see above), which needs to be investigated in detail.
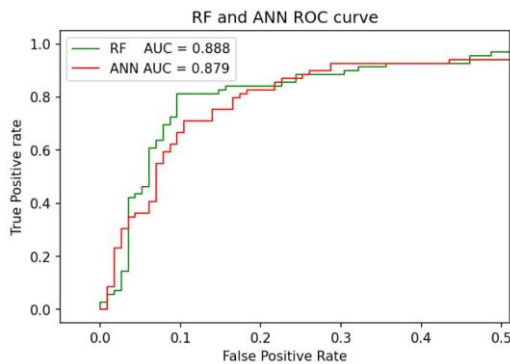

Figure 8. ROC curves obtained for the RF and for the ANN.

3.4 Feature Importance: Based on the feature importance analysis with leave-one-out approach, elevation was found significant for both models (Figure 9, Tables A1 and A2 in Appendix). However, ANN found forest type as the most important based on the F-1 score.
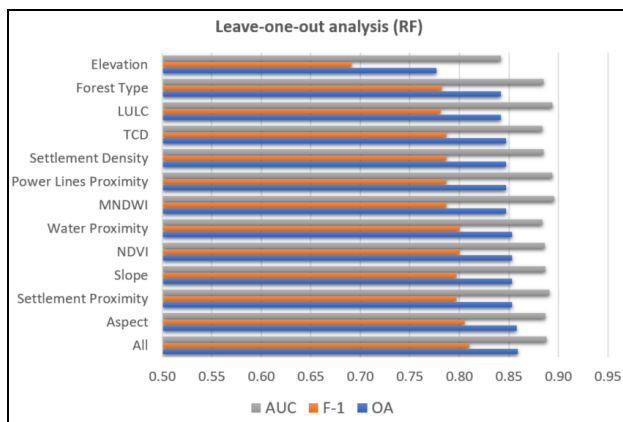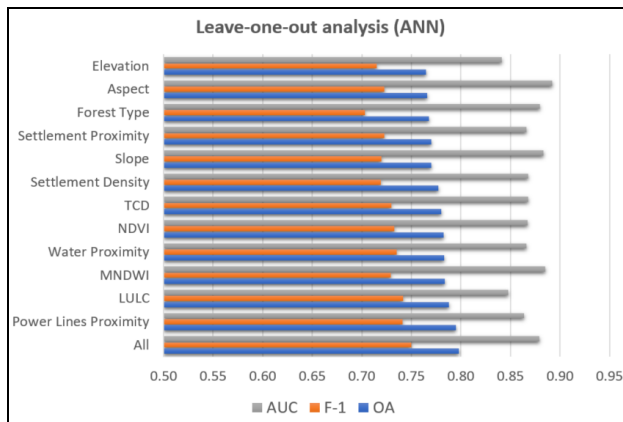

Figure 9. Leave-one-out analysis results ordered by OA values.

When the results presented here are compared with the study of Kantarcioglu et al. (2023) carried out in Thrace Region,

Türkiye, the accuracy was better in the latter one (AUC of ANN = 0.94 and RF = 0.95) possibly due to the use of extensive local database including forest roads, and the use of a larger sample size (1506). In addition, although the coverages of both sites are comparable, the environmental and anthropogenic characteristics of both regions are also different.

4. CONCLUSIONS AND FUTURE WORK

In this study, forest fire susceptibility was asessed over a region comprising four provinces in north-eastern Turkiye, using two different ML methods, namely RF and the ANN. The forest fire inventory obtained from the regional forest authority was used to identify reference ignition locations (368). The features such as elevation, land-cover, hydrology, forest, power lines were obtained from openly available datasets. The tested methods achieved comparable performances, with AUC values of 0.89 and 0.88, respectively. Yet, the ANN yielded larger areas as fire prone.

Based on leave-one-out analysis, the elevation was found to be the most influential factor. Forest type, aspect and LULC were also found highly significant, which demonstrates that vegetation characteristics, topography and anthropogenic factors need to be considered for producing accurate forest fire susceptibility maps. Yet, the results presented here are preliminary and require further investigations especially w.r.t. the different LULC types.

REFERENCES

Abadi, M., Barham, P., Chen, J.M., Chen, Z.F., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X.Q., Assoc, U., 2016. TensorFlow: A system for large-scale machine learning. In, *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)* (pp. 265-283). Savannah, GA

Adab, H., Kanniah, K.D., Solaimani, K., 2013. Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. *Natural Hazards, 65*, 1723-1743

Akinci, H.A., Akinci, H., 2023. Machine learning based forest fire susceptibility assessment of Manavgat district (Antalya), Turkey. Earth Science Informatics, 16, 397-414

Arderne, C., Zorn, C., Nicolas, C., Koks, E.E., 2020. Predictive mapping of the global power system using open data. *Scientific Data, 7*, 19

Breiman, L., 2001. Random forests. *Machine learning, 45*, 5-32

Cao, Y.X., Wang, M., Liu, K., 2017. Wildfire Susceptibility Assessment in Southern China: A Comparison of Multiple Methods. *International Journal of Disaster Risk Science, 8*, 164-181

Chicas, S.D., Østergaard Nielsen, J., 2022. Who are the actors and what are the factors that are used in models to map forest fire susceptibility? A systematic review. *Natural Hazards, 114*, 2417-2434

CLMS, 2023. European Union (EU) European Environment Agency (EEA) Copernicus Land Monitoring Service Forests. https://land.copernicus.eu/pan-european/high-resolution-layers/forests (22 February 2023)

EU-DEM v1.1, 2023. https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1 (22 February 2023).

EU-Hydro, 2023. EU-Hydro - River Network Database. https://land.copernicus.eu/imagery-in-situ/eu-hydro/eu-hydro-river-network-database (22 February 2023).

ESA WorldCover, 2023. ESA WorldCover 10 m 2020 v100. https://worldcover2020.esa.int/ (22 February 2023).

Gill, A.M., Stephens, S.L., Cary, G.J., 2013. The worldwide "wildfire" problem. *Ecological Applications, 23*, 438-454

Jain, P., Coogan, S.C.P., Subramanian, S.G., et al., 2020. A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28, 478-505, doi:10.1139/er-2020-0019.

Kantarcioglu, O., Kocaman, S., Schindler, K., 2023. Artificial Neural Networks for Assessing Forest Fire Susceptibility in Türkiye. *Ecological Informatics*, 2023. https://doi.org/10.1016/j.ecoinf.2023.102034

Leuenberger, M., Parente, J., Tonini, M., Pereira, M.G., Kanevski, M., 2018. Wildfire susceptibility mapping: Deterministic vs. stochastic approaches. *Environmental Modelling & Software, 101*, 194-203

Negnevitsky, M., 2002. Artificial Intelligence–A Guide to Intelligent Systems, Addison–Wesley Co.: Boston, MA, USA, p. 394.

OSM, 2023. OpenStreetMap Project. https://wiki.openstreetmap.org/wiki/Main_Page (22 Feb 2023)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830

Rodrigues, M., de la Riva, J., 2014. An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software, 57*, 192-201

Sachdeva, S., Bhatia, T., Verma, A.K., 2018. GIS-based evolutionary optimized Gradient Boosted Decision Trees for forest fire susceptibility mapping. *Natural Hazards, 92*, 1399-1418

World Bank, 2023. Derived Map Of Global Electricity Transmission And Distribution Lines. https://datacatalog.worldbank.org/search/dataset/0038055 (22 February 2023)

Zhang, G.L., Wang, M., Liu, K., 2021. Deep neural networks for global wildfire susceptibility modelling. *Ecological Indicators, 127*

APPENDIX

Feature Name	OA	F-1	AUC
All	0.80	0.75	0.88
LULC	0.79	0.74	0.85
Power Lines Proximity	0.80	0.74	0.86
Water Proximity	0.78	0.74	0.87
NDVI	0.78	0.73	0.87
TCD	0.78	0.73	0.87
MNDWI	0.78	0.73	0.89
Settlement Proximity	0.77	0.72	0.87
Aspect	0.77	0.72	0.89
Slope	0.77	0.72	0.88
Settlement Density	0.78	0.72	0.87
Elevation	0.77	0.72	0.84
Forest Type	0.77	0.70	0.88

Table A1. Leave-one-out results of the ANN.

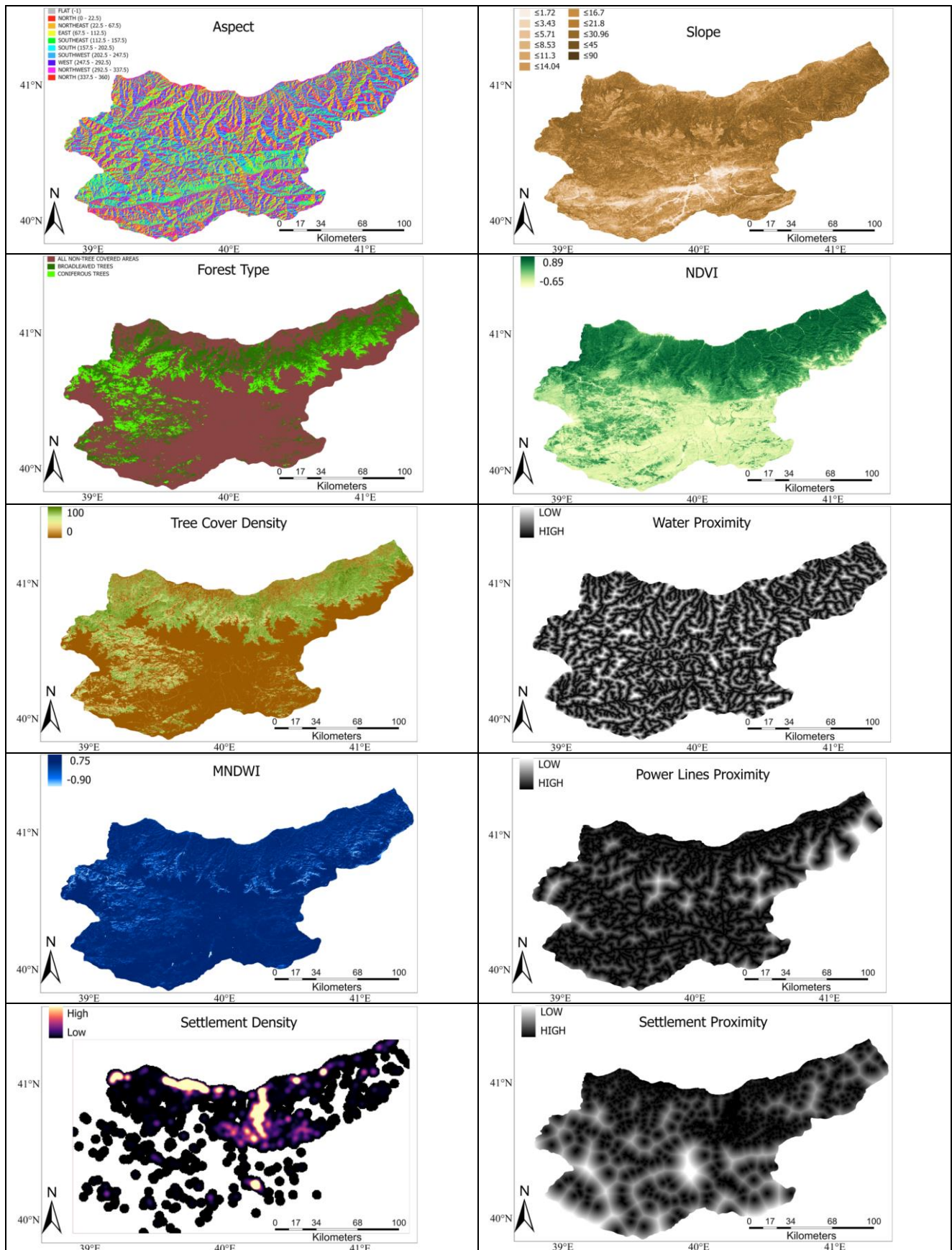| Feature Name	OA	F-1	AUC
All	0.86	0.81	0.89
Aspect	0.86	0.81	0.89
NDVI	0.85	0.80	0.89
Water Proximity	0.85	0.80	0.88
Settlement Proximity	0.85	0.80	0.89
Slope	0.85	0.80	0.89
MNDWI	0.85	0.79	0.90
Power Lines Proximity	0.85	0.79	0.89
Settlement Density	0.85	0.79	0.89
TCD	0.85	0.79	0.88
Forest Type	0.84	0.78	0.89
LULC	0.84	0.78	0.89
Elevation	0.78	0.69	0.84

Table A2. Leave-one-out results of the RF.

Figure A1. Feature maps used in the modelling.