https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

How the recursive feature elimination affects the SVM and RF for wildfire modeling? A mountainous case study area

Ali Rezaei Barzani[1], Parham Pahlavani[1], Omid Ghorbanzadeh[2], Pedram Ghamisi[2,3]

(1) School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran, ali.rezaeibarzani@ut.ac.ir;
(2) Institute of Advanced Research in Artificial Intelligence (IARAI), Landstraßer Hauptstraße 5,1030 Vienna, Austria; omid.ghorbanzadeh@iarai.ac.at; pedram.ghamisi@iarai.ac.at;
(3) Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology,09599 Freiberg, Germany
Correspondence to: pahlavani@ut.ac.ir

Abstract. This study aims to identify key fire factors via recursive feature elimination (RFE) to generate forest fire susceptibility map (FSM) using support vector machine (SVM), and random forest (RF) models. The fire zones were derived from MODIS satellite imagery from 2012 to 2017. Further validation of these data has been provided by field surveys and reviews of land records in rangelands and forests; a total of 352 fire points were determined in this study. Seventeen factors involving topography, geomorphology, meteorology, hydrology, and anthropology were identified as being effective primary factors in triggering and spreading fires in the selected mountainous case study area. As a first step, the RFE models of the RF, Extra trees, Gradient boosting, and AdaBoost were used to identify important fire factors among all selected primary factors. The SVM and RF models were applied once on all factors and the second on those derived from RFE models as the key factors in FSM. Training and testing data were divided tenfold, and the model's performance was evaluated using cross-validation (CV). Different metrics were used to measure accuracy, including recall, precision, F1, accuracy, the area under the curve (AUC), Matthews correlation coefficient (MCC), and Kappa. The accuracy assessment process shows that the FSM results are further improved by leveraging RFE models to distinguish the key factors and not include unnecessary factors. The greatest improvement is for SVM, with more than 10.97% and 8.61% in the accuracy and AUC metrics, respectively.

1. Introduction

Through the production of oxygen and the absorption of CO_2, forests play an important role in environmental pollution reduction (Pourtaghi et al., 2016). In addition to providing goods and livelihoods, forests protect plant and animal diversity, prevent soil degradation and erosion, control water flow and prevent flooding, and regulate climate by trapping carbon, which increases greenhouse gas levels (Bruinsma, 2017). Therefore, most forest fires, whether natural or induced by humans, cause many negative ecological, social, and economic impacts on forest restoration.

The term "forest fire" refers to an uncontrolled wildfire that requires some kind of action or decision to suppress. This term is used when the area of the fire exceeds 0.5 hectares and leads to the destruction of trees. Fires in forests lead to a lot of environmental destruction due to the presence of highly combustible trees (Adab et al., 2013). Forest fires are considered a widespread and critical element of the earth system (Bond and Keeley, 2005). Fire destroys nearly 350 million hectares of forest every year, according to Wright et al. (Wright et al., 2002).

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

According to Copernicus (https://effis.jrc.ec.europa.eu), fires have influenced Europe in many ways, including climate change, land management, and social patterns, such as immigration from rural settlements, rapid urbanization, and changes in leisure behavior. In recent years, forest fires have been considered one of the most critical natural hazards because they cause loss of life, severe infrastructure damage, and life-threatening environmental effects (Tien Bui et al., 2016).

In recent years, fires in forested areas in northern Iran have become a major threat (Adab et al., 2015). Iran's Forest Organization estimates that 5,000 to 6,000 hectares of forest land are destroyed yearly due to frequent fires (Adab et al., 2013). Natural processes have historically caused fires in forests. However, the firing process has accelerated due to increased human-environmental interactions. This makes it more likely to occur in the future (Valdez et al., 2017). Many factors, such as wind, topography, and drought, play an essential role in the occurrence and spread of fires, but in many cases, human factor causes fires in forests (Sayad et al., 2019). The situation is also the same for the occurrence of this hazard in northern regions of Iran. Therefore, the forest fires in this area are mainly due to human activities from recreational camps and land clearing to change land use and remove livestock grazing from rangelands and forests (Jahdi et al., 2014). It is important to take precautions and predict fire facilities in fire-prone areas to prevent forest fires. Therefore, preparing a fire susceptibility map (FSM) and identifying hazardous and fire-prone areas are essential (Eskandari et al., 2021).

When preparing a susceptibility map, it is important to identify factors that contribute to the fire's susceptibility. Geospatial predictive models require data from remote sensing and geographic information system (GIS) sources. Further, FSM must be produced based on appropriate input data and an appropriate methodology (Jaafari and Pourghasemi, 2019). This challenge has opened a way for research and study in this field to create a more effective and efficient understanding of these geo-environmental factors by analyzing the data and leading to better management in this area. In any natural hazard susceptibility modeling and for the forest fire, it should be examined whether variables associated with forest fires can be used for event prediction (Hong et al., 2018; Jaafari and Pourghasemi, 2019). In addition, variables with low prediction capability should be excluded from predictive models to increase accuracy. The most important and optimal number of fire-predictive variables must be identified to create high-reliability fire susceptibility maps (Hong et al., 2018).

Many studies have examined fire susceptibility modeling and FSM in forest areas around the world. The spatial pattern of wildfire probability has been simulated and predicted using different spatial modeling strategies across different geographical regions. Several studies have used the SVM and RF models to prepare fire risk maps. Jaafari et al. reported (Jaafari and Pourghasemi, 2019) an AUC value of 0.75% for the SVM model with nine factors in the Chaharmahal Bakhtiari Province in Iran. Hong et al. (Hong et al., 2018) evaluated two machine learning (ML) models, including the RF and SVM models with eight factors in China, and achieved an AUC value of 84% and 74% for the RF and SVM, respectively. RF and SVM are successful ML models in fire susceptibility mapping (Kalantar et al., 2020; Tavakkoli Piralilou et al., 2022). Yousefi et al. (Yousefi et al., 2020) used three ML models to produce a multi-hazard risk map for a region in Iran. For producing a wildfire susceptibility map, twelve factors were considered. The AUC value of 83% was obtained for the SVM model. Tonini et al. (Tonini et al., 2020) produced a fire susceptibility map for two seasons using the RF model. For validation, they used k-fold cross-validation. Ghorbanzadeh et al. (Ghorbanzadeh et al., 2019) examined the performance of ANN, SVM, and RF ML models with seventeen factors to predict wildfire-prone areas in Mazandaran Province, Iran, and

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

reported AUC values of 88% and 78% for the RF and SVM models, respectively. The literature review clearly shows that most methodologies employed conditional wildfire criteria without any evaluation if they play a significant role in their selected case study area.

Forest areas vary in topography, climate zone, settlement density, human activity, etc. There is, however, a significant gap in the literature. While some studies have used several fire conditional factors, it is unclear whether the limited derived accuracies of the resulting FSMs caused by the model limitations or the adverse impact of some not fully related factors. The methods used to select optimal inputs are known as feature selection techniques. Feature selection may be categorized into filter and wrapper methods and embedded methods (Kohavi, 1997). Filter methods are used to identify the importance of factors (Guo et al., 2016; He et al., 2021; Jaafari et al., 2018; Ngoc Thach et al., 2018; Tehrany et al., 2019; Tien Bui et al., 2016) and rely on the characteristics of the data. These methods do not use ML algorithms. To identify the optimal number of variables predicting forest fires, Hong et al. (Hong et al., 2018) used genetic algorithms, a metaheuristic method. After determining the optimal number of variables, ML methods were used to prepare fire susceptibility maps. Jafari and Pourghasemi first identified important factors using the RF model and then applied the SVM model to obtain the fire risk map (Jaafari and Pourghasemi, 2019). Also, Eskandari, in his paper, first identified important factors using the RF model and then applied his models to these factors (Eskandari et al., 2021). Pourghasemi et al. have used the Boruta algorithm to identify the factors affecting forest fires and floods (Pourghasemi et al., 2020). Boruta algorithm has a behavior similar to the RF model in identifying the importance of variables. These methods only compare the correlation between factors, and this causes the impact of several characteristics to be ignored by putting together the occurrence of fire. However, the wrapper methods cover this defect and consider the combination of features as effective features in the problem. The wrapper method is also called the greedy search algorithm because this method scans all possible combinations of features before selecting the one that produces the best ML algorithm performance. The embedded methods consider the interaction between features and models as part of the feature selection process (Agrawal et al., 2021). Computationally, the embedded methods cost less than wrapper methods because an analysis for the optimum feature subset and formulation of a model in an embedded method can be performed simultaneously (Remeseiro and Bolon-Canedo, 2019). Some examples of embedded methods include decision tree-based algorithms such as a random forest. Recursive feature elimination (RFE) is a hybrid between embedded and wrapper methods (Guyon et al., 2002).

In this study, the RFE method was used to select features. Features are conditional wildfire criteria. As a result of this method, the predictors are first created into a model, which is then used to calculate an importance score for each predictor based on the model. In RFE, the AdaBoost and Gradient Boosting, Random Forests, and Extra trees were used. After that, features that appear in more than two models were voted on and selected as the final features. Therefore, the objectives of this study are to present an effective feature selection method for identifying effective factors involved in forest fires, testing the proposed feature selection method against the SVM and RF models in order to determine the improvement in accuracy and improved mapping of forest fire-prone areas in the study area.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

2. Study Area

The study area is Amol County in the Mazandaran province in northern Iran. This Province, with an area of approximately 23,842 km^2 lies on the southern shore of the Caspian Sea (see Figure 1). The Amol forest area lies in the center of Mazandaran province, in the city of Amol. This study area is mostly mountainous and elevated and covers approximately 646 square kilometers. This Province is known to be one of the most wildfire-prone regions in northern Iran (Adab et al., 2015). The minimum and maximum altitudes in the region were 100 and 2500 m above the mean sea level. Every year, tourists visit this region, and tourism is one of the residents' primary income sources (Ghorbanzadeh et al., 2019). Therefore, protecting and maintaining forest areas is also important economically apart from the environmental discussion.

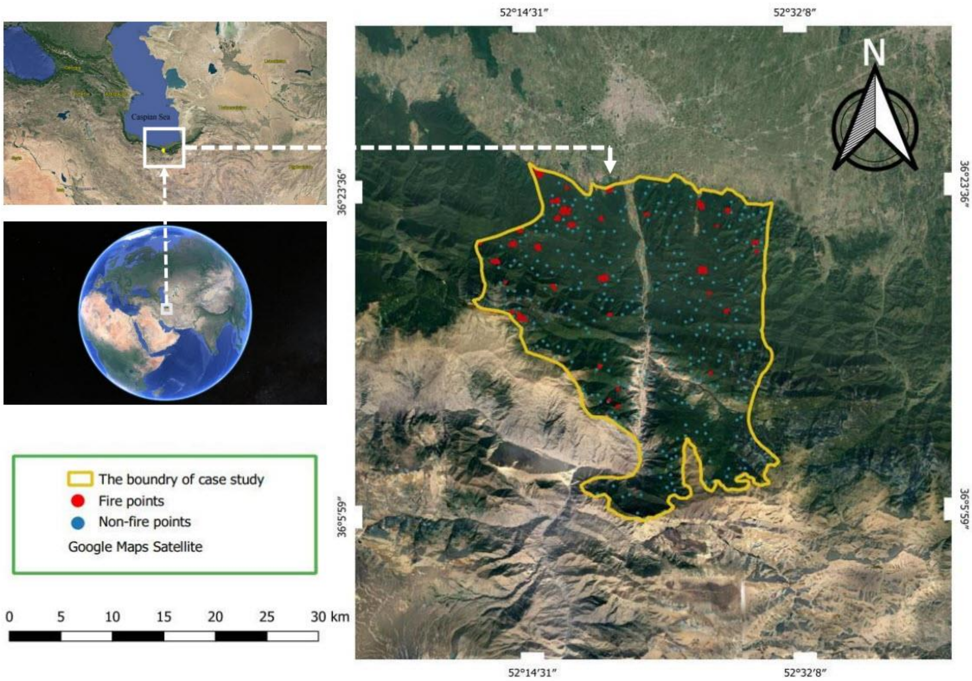


Figure 1. The location of the study area. The figure was drawing based on © Google Earth 2022 (https://earth.google.com).

3. Materials

3.1. Forest fire inventory data set for training and testing the models

The first step in producing an FSM is to prepare an inventory dataset of historical forest fire events in the study area. This study used recorded reports, satellite imagery, and field survey to map the fires. The MODIS fire data were used that are available for free from https://modis.gsfc.nasa.gov/data/. The MODIS fire inventory dataset used in the analysis ranges from 2012 to 2017. Field surveys and forest organization reports verified the fire inventory dataset. Accordingly, 352 fire points were identified in the study area. In this study area, 34 polygons

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

of fire have been identified according to previous research. However, in this study, fire points inside each polygon were randomly extracted according to the need for fire points. Also, it was set to consider a minimum distance of 50 meters between the two fire points. Of the 34 polygon fires, 352 fire points were extracted. For non-fire points, in areas where the fire did not occur, 352 non-fire points were randomly extracted to maintain the balance between the two classes of fire and non-fire. In non-fire zones, it was set not to have a distance of fewer than 50 meters to see a suitable distribution of points in the study area. This study used the k-fold method to test and validate the data. In this method, the datasets are divided into k-fold. At each stage of the CV process, the model is trained on k-1 folds and tested with a residual fold. This process is repeated k times, and finally, the average evaluation scores obtained from each iteration are calculated. In this study, the most common number of k=10 was followed according to the literature review (Kalantar et al., 2020; Tavakkoli Piralilou et al., 2022).

3.2. Forest fire conditional factors

In order to produce FSM, 17 important factors were extracted. Factors can be classified into five categories: anthropological, topographical, vegetation, meteorological and hydrological. Anthropological factors are distance to villages, distance to the recreation area, distance to road, distance to power transmission lines, distance to mines, and Land-use. The first six factors were obtained through the state wildlife organization of Amol County (SWOAC), and the Land-use factor was obtained from Landsat satellite images. The vegetation factor in this study is the NDVI factor obtained from Landsat 8 satellite images. The extracted topography factors are the altitude, slope, aspect, landforms, topographic wetness index (TWI), and plane curvature. All of these topographic factors have been extracted from Aster Dem. Two meteorological factors used in this study are annual temperature and the wind effect obtained from Amol County (SMOC) state organization for the study area. Hydrological factors include annual rainfall and distance to the river extracted from SMOC.

Table 1. Explanatory factors were used in this study.

No	Conditioning Factor	Type	Source
1	Altitude (m)	Continuous	ASTER DEM
2	Annual temperature (◦C)	Continuous	SMOAC
3	Annual rainfall (mm)	Continuous	SMOAC
4	Topographic wetness index (TWI)	Continuous	ASTER DEM
5	Landform	Categorical	ASTER DEM
6	Land use	Categorical	LANDSAT 8
7	NDVI	Continuous	LANDSAT 8
8	Distance to villages (m)	Continuous	SWOAC
9	Distance to recreation area (m)	Continuous	SWOAC
10	Distance to river	Continuous	SWOAC
11	Distance to road	Continuous	SWOAC
12	Distance to mine	Continuous	SWOAC

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

13	Distance to power transmission lines	Continuous	SWOAC
14	Slope	Continuous	ASTER DEM
15	Aspect	Categorical	ASTER DEM
16	Plan curvature (100/m)	Categorical	ASTER DEM
17	Wind effect	Continuous	SMOAC

4. Methods

The research process is divided into five steps, the second and third of which are devoted to our proposed approach for identifying the factors affecting forest fires:

- identifying the forest fire factors associated with the study area,
- developing the RFE method for choosing features according to four different models,
- voting for the best features and selecting the ultimate ones,
- using two ML models of the SVM and RF for FSM based on all factors. Then taking into account the superior factors, validate the resultant FSMs, and monitor the accuracy improvements.

The flowchart of the method used in the research is shown in Figure 2.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

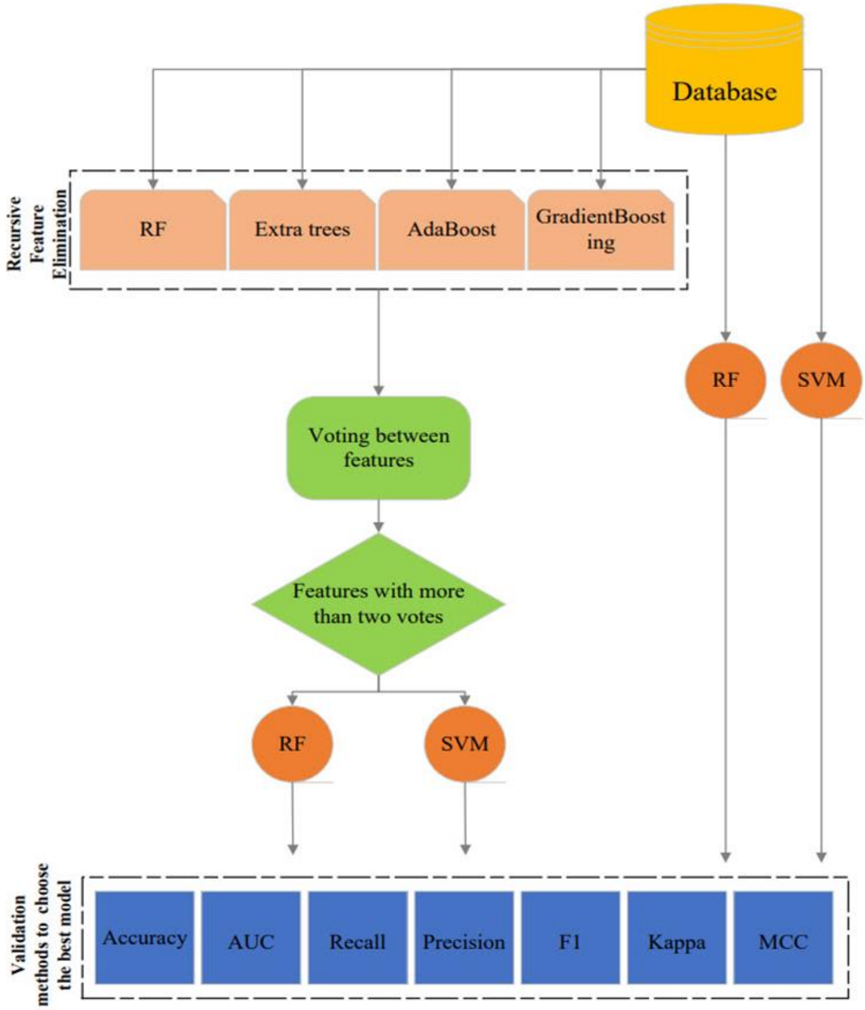


Figure 2. The flowchart signifies this paper's introduced approach for forest fire susceptibility modeling and mapping.

4.1. Recursive Feature Elimination

Feature selection is accomplished by recursive feature elimination (RFE), one of the most popular methods of selecting features developed by Guyon et al. (Guyon et al., 2002). As a result of its simplicity and effectiveness, RFE is a popular feature selection algorithm because it can identify the feature in a training dataset that is most relevant to the prediction of the variable. In this method, the properties are selected recursively and take into account the smaller sets of features at each stage (Kuhn and Johnson, 2013). The properties in RFE are ranked based on the order of their removal from the property space. Feature selection is supervised in this method and requires labeled data for the best performance. In RFE, features are gradually removed so that only useful features remain, using the output function to determine which features are relevant and which are not. By building a model

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

on the entire set of problem variables, this method calculates their importance for each variable and then reverses the process to select variables inversely (Guyon et al., 2002). In the next step, the variable is removed with the lowest importance, the model is re-constructed on other remaining variables, and the importance of each variable is recalculated. Until the targeted variable number is reached, the trend continues. There is a limitation on the number of features that can be kept in the RFE, but the number of valid features is usually not known in advance. In this study, cross-validation is used to find the optimal number of features for scoring different feature subsets and choosing features with the highest score. When the full model is created, a measure of variable importance is computed that ranks the predictors from the most important to the least. Models based on linear models have coefficients, whereas models based on decision trees have feature importance. Therefore, the estimator is trained, and the features are chosen based on the coefficients or on the importance of the features. Next, the least important feature is removed, and a new ML model is built utilizing the remaining features. To determine the performance of this model using accuracy assessment metrics, the performance against the full model is compared. This procedure is continued over and over through all applied features in our training dataset. Using it, it could be possible to scan all the features in the dataset to determine which ones contributed to a significant drop in performance and should be kept and which ones should be removed. Here, the applied ML models in the RFE method are introduced and detailed.

4.2. Random Forest

The RF model was created by Ho (1994) and developed by Breiman (Breiman, 2001). The RF model consists of several decision trees. Instead of relying on a decision tree, RF takes the prediction from each tree and predicts the final output based on the majority of votes. It enhances the accuracy of the model and prevents the overfitting issue. The RF model is also used in regression problems in addition to classification problems. The RF hyperparameters must be adjusted before training to get a good result from the model. An evaluation of a variable's importance in RF models is based on the *Gini* index (Cao et al., 2017). In this study, the grid search method (Ngoc Thach et al., 2018) was used to determine the optimum value for model hyperparameters. The grid search method obtains the maximum number of trees and the number of variables. The maximum number of variables was set on the squared root of the total number of variables in both 8 and 17 variable models. A grid search method was used to obtain the maximum of 1000 trees in the 8-variable model and 200 trees in the 17-variable RF model.

4.3. Extra trees

The Extra trees is an Ensemble algorithm invented by Geurts et al. (Geurts et al., 2006). This algorithm combines the predictions of several decision trees to make the final prediction (Geurts et al., 2006). Each decision tree fits the entire training dataset in this algorithm, as opposed to the RF, which only fits a bootstrap sample of the dataset. According to the classic decision tree procedure, the algorithm makes a set of decision trees from top to bottom (Ampomah et al., 2020). The final prediction is based on a majority vote that runs between the predictions of all trees. The two main differences between this method and tree-based ensemble methods are that it splits nodes by selecting completely random cut-points, and as the second difference, it uses a whole sample of the training dataset instead of bootstrap samples (Geurts et al., 2006). In the RF model, the cut-off points are selected optimally in order to split nodes, but in this method, these points are selected completely randomly. This randomness of the cut-off points leads to a decrease in variance. Using all dataset training samples instead of bootstrap samples

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

reduces the bias of the problem (Ampomah et al., 2020). In this study, hyperparameters were set to 100 for the number of trees and to the rounded-up square root of the number of predictor factors for the maximum feature.

4.4. AdaBoost Model

The AdaBoost algorithm was proposed by Freund and Schapire in 1997 (Freund and Schapire, 1997). An ensemble algorithm creates a strong classifier from a set of weak classifiers (Devi et al., 2020). By combining the weak classifiers, AdaBoost can adjust the weak errors and form a stronger final classifier. As a result, the accuracy of the strong classifier is determined by the accuracy of the weak classifier. The process reduces bias and variance, thereby improving classification ability and efficiency (Wu et al., 2020). The accuracy of the final learner may be affected by outliers in AdaBoost (He et al., 2021). The iteration number was set to 100 in this study, and the learning rate was set to 1. We also weighed the weak classifier based on the classification effect of the sample set

4.5. Gradient Boosting Decision Tree

The Gradient boosting decision tree (GBDT) was proposed by Fridman in 2001, which is an approximation method using the gradient descent method (Song et al., 2018). The GBDT is an ensemble machine-learning method combining multiple decision trees based on the boosting concept (He et al., 2021). It models the data with a decision tree (DT) as the basic classifier and Gradient boosting as the training strategy (Liang et al., 2021). Whereas random forests build an ensemble of deep independent trees, GBDT builds an ensemble of shallow trees in sequence, with each tree learning and improving on the previous one. The GBDT is capable of constructing a strong classification model by combining weak classifiers after multiple iterations. By iterating, the previous model's residuals are reduced, and the results are improved (He et al., 2021). Like the previously applied machine learning models in RFE, the maximum number of iterations was set to 100, and the learning rate is considered 0.1.

4.6. Voting

The next step is to vote on the superior features identified by each model and the features identified as superior by more than three models. Following that, two ML models of the SVM and RF have been used in the preparation of FSMs in this study. In this stage, these two models were trained on all factors once. Furthermore, the second time, only the superior factors obtained from the RFE method were used, and the resulting FSMs were validated to determine the extent of improvement.

4.7. Support Vector Machine

The SVM is one of the most widely used and successful ML techniques for classification and regression (Vapnik, 1998). The type of kernel function used in the SVM model plays an important role in its performance. In modeling the natural hazard susceptibility, using SVM with radial basis function kernel is common. This method is one of the most common approaches in the field of natural disasters that used in this study. The performance of the SVM model is influenced by the two hyperparameters. These two hyperparameters are kernel width (γ) and regularization (C). Kernel width defines how far it influences the calculation of a plausible line of separation. When hyper-parameter γ is higher, nearby samples will have a high influence, and low γ means far away samples

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

also be considered to get the decision boundary. The C tells the SVM optimization how much error is bearable. When C is high, it will classify all samples correctly; also, there is a chance to overfit. In this study, the grid search method was used to determine the desired values for C and γ. For SVM model with 17 factors, C = 1 and γ = 0.1 and for SVM model with 8 factors, C = 1 and γ = 1.

5. Experimental results

In the following, the results of the implemented methods are given along with figures and diagrams, and the description and details of these results are discussed. All of the mentioned data were processed and used to create the FSM-related factors with ArcGIS 10.8. All methods were implemented by sklearn (Pedregosa et al., 2012), which is an open-source Python library, in the google colab environment.

5.1. Feature selection using the RFE method

In Figures 3, 4, 5, and 6, the Y-axis indicates the accuracy of the model, and the X-axis shows the number of features on which the model is prepared. It is known that the dashed line is drawn where the highest accuracy has been achieved, and the number of features achieved is determined. Figure 3 shows that the RF model's highest accuracy was obtained when eight features were included. Figure 4 specifies that the AdaBoost model's highest accuracy occurred with ten features. Moreover, in Figure 5, it is clear that the Gradient Boosting model obtained its highest accuracy with eight characteristics. Figure 6 shows that the Extra Trees model achieved its highest accuracy with eight features. Accordingly, Table 2 lists all factors selected or rejected in four models based on the Recursive Feature Elimination (RFE) methodology.

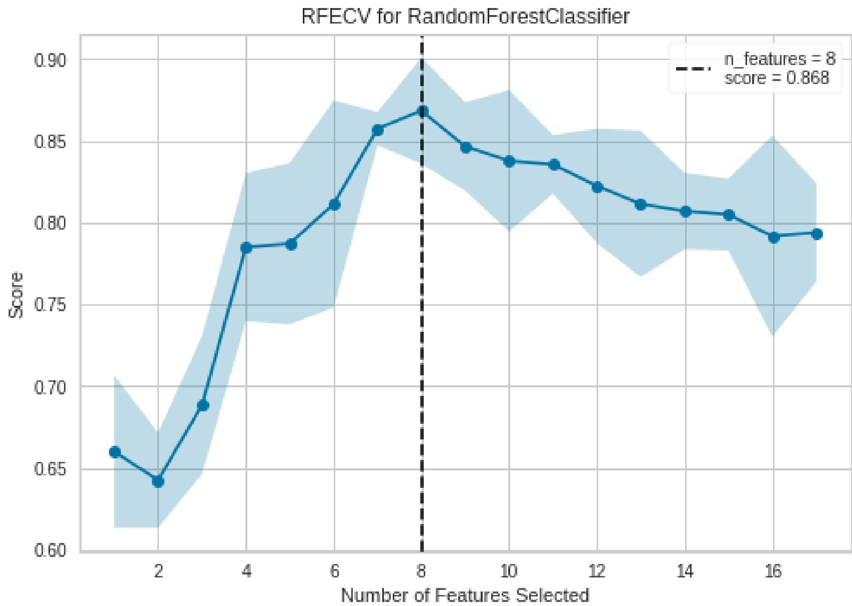


Figure 3. Recursive feature elimination with cross-validation (RFECV) for the RF model.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

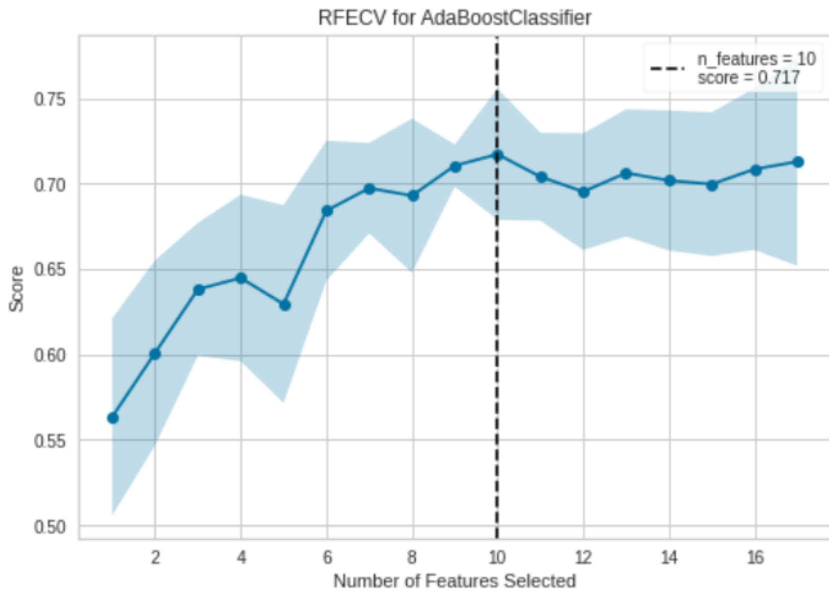


Figure 4. Recursive feature elimination with cross-validation (RFECV) for the AdaBoost model.

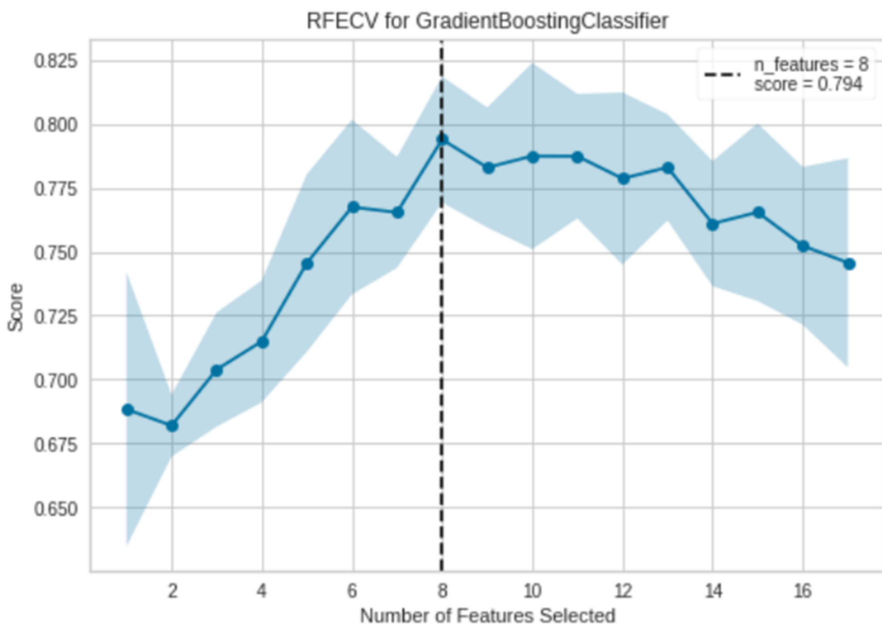


Figure 5. Recursive feature elimination with cross-validation (RFECV) for the GBDT model.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

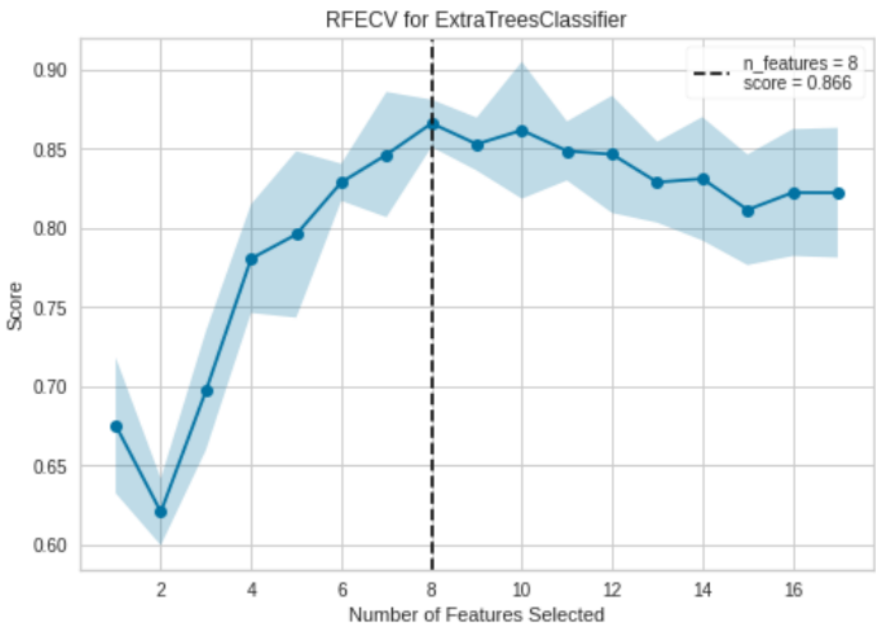


Figure 6. Recursive feature elimination with cross-validation (RFECV) for the extra trees model.

Table 2. Considering forest fire variables importance using recursive feature elimination (RFE).

Factors	Random Forest	AdaBoost	Gradient Boosting	Extra trees	Voting
Slope	Rejected	Rejected	Rejected	Rejected	Removed
Altitude	Confirmed	Rejected	Confirmed	Confirmed	Selected
Aspect	Confirmed	Confirmed	Confirmed	Confirmed	Selected
Plan curvature	Rejected	Rejected	Rejected	Rejected	Removed
TWI	Rejected	Confirmed	Rejected	Rejected	Removed
Distance to river	Confirmed	Confirmed	Confirmed	Confirmed	Selected
Distance to road	Confirmed	Confirmed	Confirmed	Confirmed	Selected
Distance to recreation area	Rejected	Confirmed	Rejected	Rejected	Removed
Distance to power lines	Confirmed	Rejected	Confirmed	Confirmed	Selected
Distance to mine	Rejected	Confirmed	Rejected	Rejected	Removed
Rainfall	Confirmed	Confirmed	Confirmed	Confirmed	Selected
Temperature	Confirmed	Confirmed	Confirmed	Confirmed	Selected
Distance to villages	Confirmed	Confirmed	Confirmed	Confirmed	Selected
Wind effect	Rejected	Rejected	Rejected	Rejected	Removed
NDVI	Rejected	Confirmed	Rejected	Rejected	Removed
Landform	Rejected	Rejected	Rejected	Rejected	Removed
Land use	Rejected	Rejected	Rejected	Rejected	Removed

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

5.2. Voting results of forest fire factors

In Figure 7, the results of voting between models are displayed as bar charts. Each factor votes for the number of models that identified that factor as an effective factor. Based on this diagram, six factors have six votes, i.e., all four models identified these six factors as effective. Figure 7 shows our final list of factors that have been identified by more than three models.

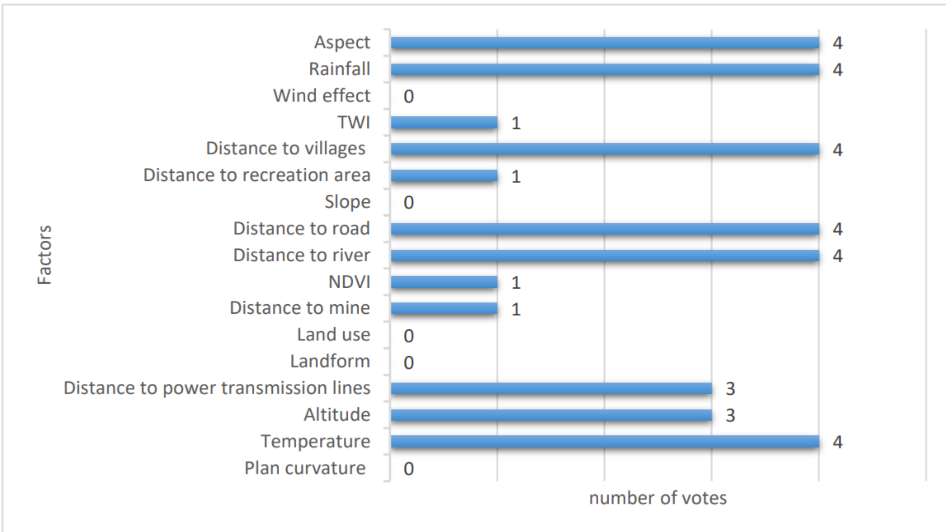


Figure 7. Votes of each factor.

5.3. Model Accuracy

In this study, several model validation methods were used: AUC, accuracy, recall, precision, F1, Kappa, and MCC. The ROC curve was plotted with the true positive, which represents a correctly predicted forest fire on the X-axis, and the false positive, which represents a falsely predicted forest fire on the Y-axis, as inputs. The area under the ROC curve represents AUC. Accuracy, recall, precision, F1, MCC, and Kappa criteria were calculated from Eq. (1-6) as follows (Chicco and Jurman, 2020; Pham et al., 2020):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \quad (5)$$

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

$$Kappa = \frac{(TP + TN) - ((TP + FN)(TP + FP) + (FP + TN)(FN + TN))}{1 - ((TP + FN)(TP + FP) + (FP + TN)(FN + TN))} \tag{6}$$

where true positive (*TP*) is the number of forest fire points categorized correctly as forest fire and true negative (*TN*) denotes the number of non-forest fire points correctly classified as non-forest fire points. Meanwhile, false positives (*FP*) and false negatives (*FN*) refer to the number of forest fire points that were incorrectly classified as forest fire or non-forest fire points. In the form of four tables, the evaluation results of each model are displayed according to the number of factors considered in each model. Tables 3 and 5 are related to RF and SVM models in which 17 primary factors are involved, Tables 4 and 6 are related to RF and SVM models with the final eight factors selected, and the factors referred to in Table 2 are involved.

Table3 The accuracy assessment results for each fold and the CV values for RF with all factors.

Fold	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
1	0.8261	0.8987	0.7727	0.8500	0.8095	0.6502	0.6527
2	0.8913	0.9545	0.8636	0.9048	0.8837	0.7818	0.7825
3	0.8478	0.9195	0.9091	0.8000	0.8511	0.6968	0.7028
4	0.8261	0.9356	0.8182	0.8182	0.8182	0.6515	0.6515
5	0.8696	0.9318	0.9091	0.8333	0.8696	0.7396	0.7424
6	0.9348	0.9924	1.0000	0.8800	0.9362	0.8701	0.8775
7	0.8889	0.9980	1.0000	0.8148	0.8980	0.7788	0.7985
8	0.8889	0.9239	0.9545	0.8400	0.8936	0.7783	0.7853
9	0.9333	0.9545	0.8636	1.0000	0.9268	0.8662	0.8741
10	0.9556	0.9625	0.9545	0.9545	0.9545	0.9111	0.9111
Mean	0.8862	0.9472	0.9045	0.8696	0.8841	0.7724	0.7778
Std	0.0429	0.0300	0.0717	0.0622	0.0458	0.0861	0.0872

Table 4. The accuracy assessment results for each fold and the CV values for RF with eight final factors.

Fold	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
1	0.9565	0.9714	0.9048	1.0000	0.9500	0.9117	0.9153
2	0.8913	0.9545	0.9091	0.8696	0.8889	0.7826	0.7833
3	0.7826	0.9233	0.9545	0.7000	0.8077	0.5709	0.6078
4	0.9130	0.9886	0.9545	0.8750	0.9130	0.8264	0.8295
5	0.9783	0.9943	1.0000	0.9565	0.9778	0.9565	0.9574
6	0.8696	0.9413	0.9091	0.8333	0.8696	0.7396	0.7424
7	0.8444	0.9028	0.8095	0.8500	0.8293	0.6866	0.6873
8	0.9333	0.9782	0.9048	0.9500	0.9268	0.8657	0.8665
9	0.8889	0.9425	0.8095	0.9444	0.8718	0.7748	0.7819
10	0.9111	0.9821	0.9524	0.8696	0.9091	0.8225	0.8257
Mean	0.8969	0.9579	0.9108	0.8848	0.8944	0.7937	0.7997
Std	0.0535	0.0287	0.0584	0.0805	0.0496	0.1058	0.0988

Table 5. The accuracy assessment results for each fold and the CV values for SVM with all factors.

Fold	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
1	0.7609	0.7879	0.7273	0.7619	0.7442	0.5199	0.5204
2	0.8478	0.8523	0.8182	0.8571	0.8372	0.6945	0.6952
3	0.7826	0.8030	0.8636	0.7308	0.7917	0.5677	0.5764
4	0.7826	0.8769	0.8636	0.7308	0.7917	0.5677	0.5764
5	0.7609	0.8598	0.7273	0.7619	0.7442	0.5199	0.5204
6	0.8043	0.9375	0.9091	0.7407	0.8163	0.6116	0.6264
7	0.7778	0.8933	1.0000	0.6875	0.8148	0.5597	0.6234

8	0.7778	0.8419	0.9091	0.7143	0.8000	0.5580	0.5787
9	0.8222	0.8518	0.7727	0.8500	0.8095	0.6436	0.6461
10	0.8667	0.9545	0.9091	0.8333	0.8696	0.7337	0.7366
Mean	0.7984	0.8659	0.8500	0.7668	0.8019	0.5976	0.6100
Std	0.0345	0.0499	0.0839	0.0565	0.0362	0.0686	0.0666

Table 6. The accuracy assessment results for each fold and the CV values for SVM with eight final factors.

Fold	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC
1	0.8478	0.9238	0.9048	0.7917	0.8444	0.6968	0.7028
2	0.9565	0.9527	0.9545	0.9545	0.9545	0.9129	0.9129
3	0.8043	0.9091	0.9545	0.7241	0.8235	0.6131	0.6429
4	0.8913	0.9413	0.8636	0.9048	0.8837	0.7818	0.7825
5	0.9783	0.9943	1.0000	0.9565	0.9778	0.9565	0.9574
6	0.8261	0.9242	0.8182	0.8182	0.8182	0.6515	0.6515
7	0.8667	0.9107	0.8095	0.8947	0.8500	0.7305	0.7335
8	0.8667	0.9147	0.8571	0.8571	0.8571	0.7321	0.7321
9	0.8889	0.9405	0.8095	0.9444	0.8718	0.7748	0.7819
10	0.9333	0.9940	0.9524	0.9091	0.9302	0.8665	0.8673
Mean	0.8860	0.9405	0.8924	0.8755	0.8811	0.7717	0.7765
Std	0.0531	0.0300	0.0666	0.0731	0.0524	0.1057	0.1012

5.4. Model Comparison

In Figure 8, the evaluation criteria for the SVM model with all factors and SVM with eight factors are displayed simultaneously to create a visual comparison. According to Figures 8 and 9, 8 selected factors from the voting process have improved all evaluation criteria in both models. The SVM model has a higher rate of improvement than the RF model. Also, according to the results of the RF model, it has shown better performance than the SVM model and has been able to achieve nearly 80% and above in all evaluation criteria. In most studies, the AUC criteria for evaluating models are reported; both models have been improved in this study. In the AUC criteria, in the SVM-8 factors model, 8.61% compared to the SVM-All factors model and 1.13% improvement in the RF-8 factors model compared to the RF-All factors model. However, since this criterion generally does not represent the performance of the models, in addition to AUC, six other evaluation criteria have been reported for the models. The SVM model's lowest improvement in evaluation criteria is related to the Recall criterion, which has improved by 4.98%, and the highest improvement rate is related to Kappa with 29.13%. In the RF model, the highest rate of improvement in the MCC criterion and the lowest improvement rate for the Recall criterion were 2.81% and 0.69%, respectively.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

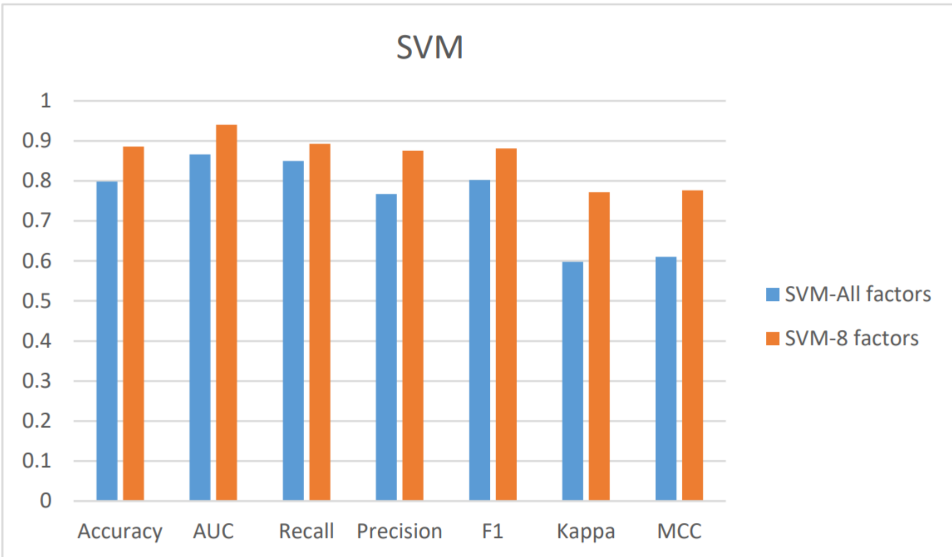


Figure 8. The resulting accuracy metrics for SVM.

In Figure 9, the evaluation criteria for the RF model with all eight factors are displayed simultaneously to create a visual comparison. In Figure 10, all models are displayed simultaneously in the same form for all eight factors that, indicate the superiority of the RF model over the SVM model. 330

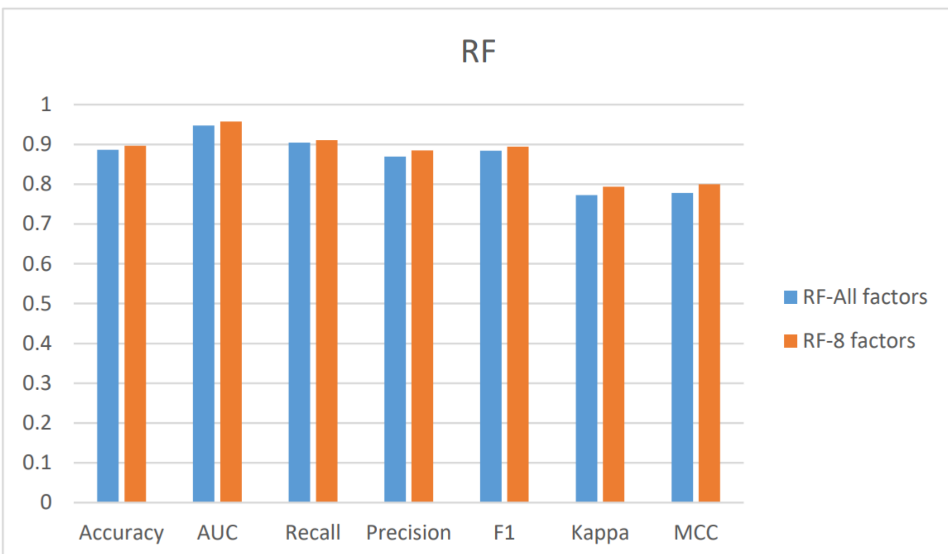
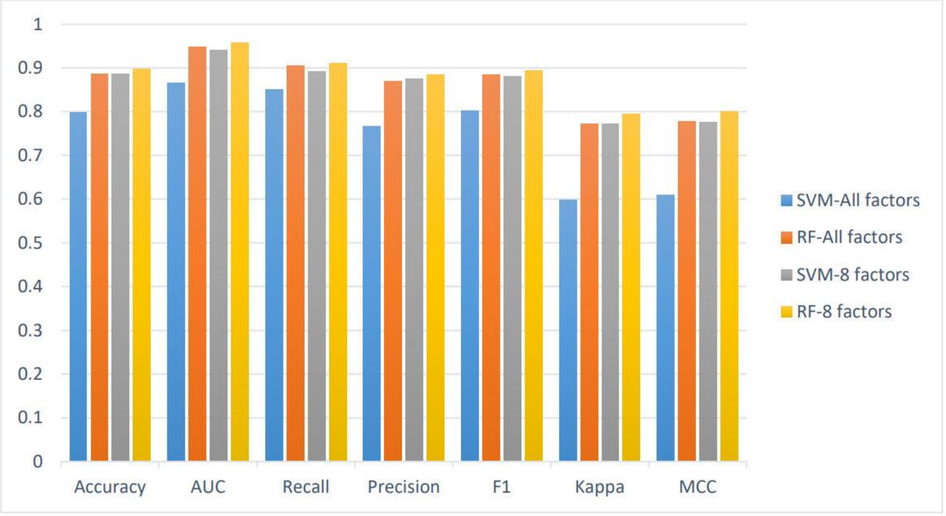


Figure 9. The resulting accuracy metrics for RF.



335

Figure 10. The resulting accuracy metrics for RF and SVM.

5.5. Forest fire susceptibility map (FSM)

The fire susceptibility map was prepared in 5 classes (Figures 11, 12, 13, and 14). In Table 7, the area of each class is specified.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

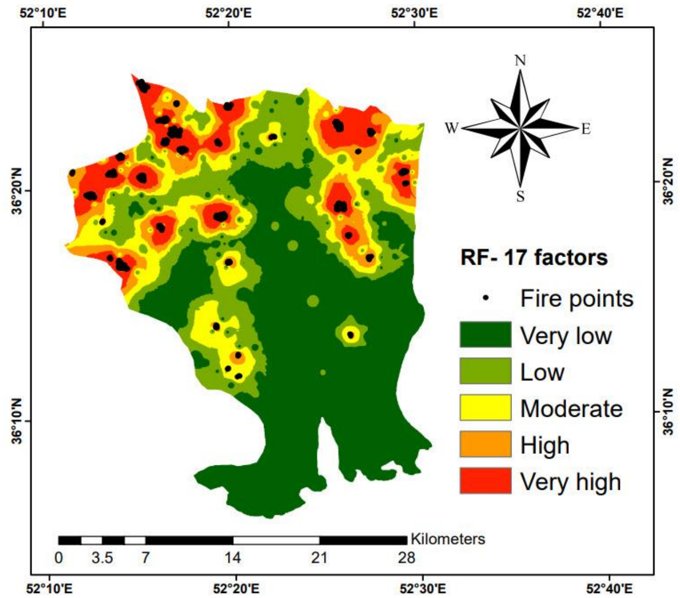


Figure 11. Forest fire susceptibility map for the RF model with all factors.

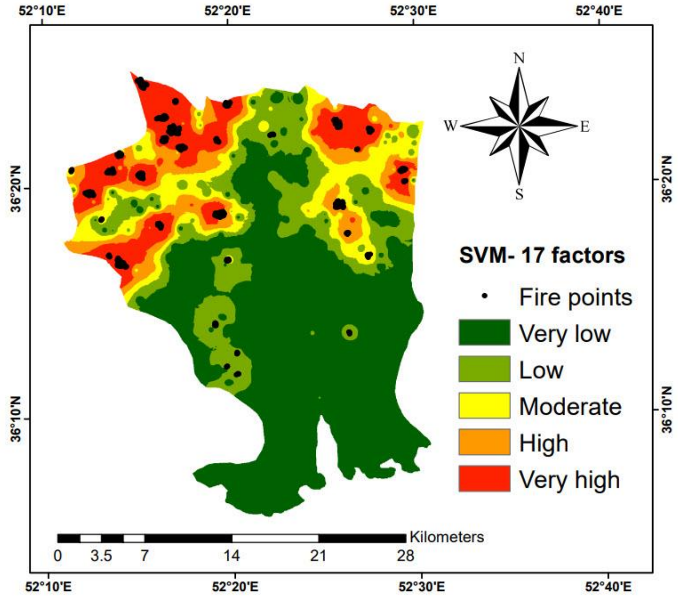


Figure 12. Forest fire susceptibility map for the SVM model with all factors.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

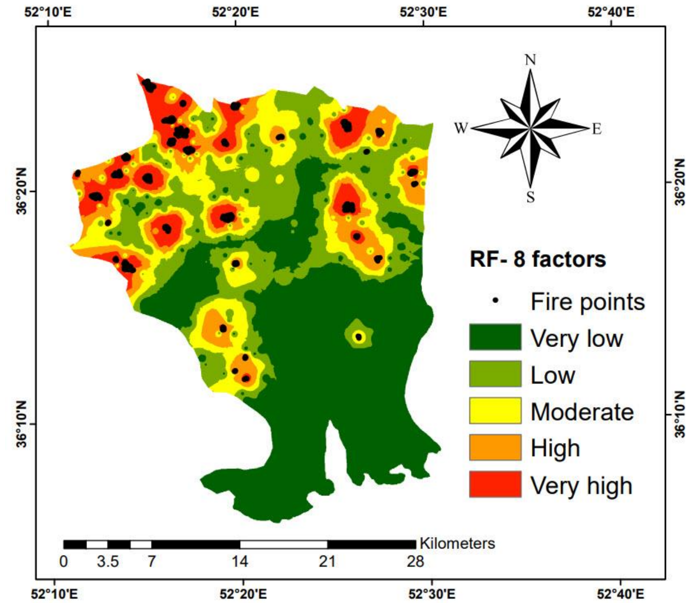


Figure 13. Forest fire susceptibility map for the RF model with eight factors.

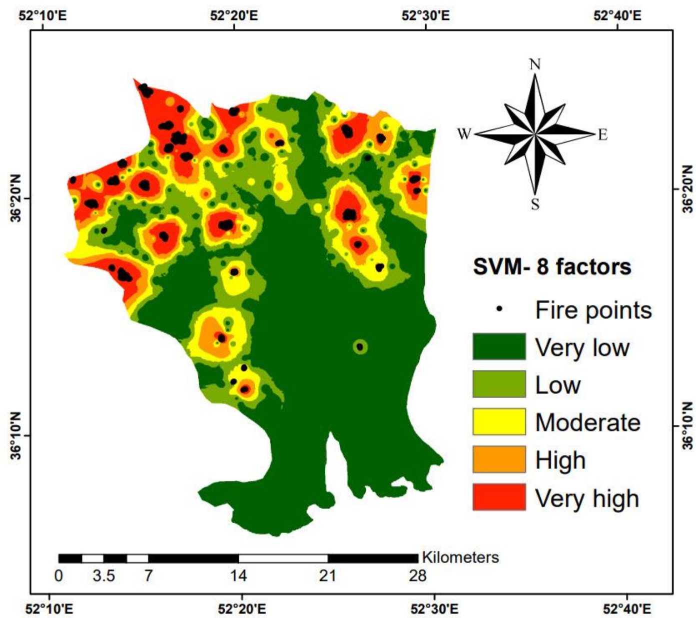


Figure 14. Forest fire susceptibility map for the SVM model with eight factors.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

Table 7. Area of each class in percentage.

Class	RF-all factors	SVM-all factors	RF-8 factors	SVM-8 factors
Very Low	46.41	51.13	43.06	53.41
Low	20.55	16.98	22.78	16.99
Moderate	13.18	10.28	14.31	11.22
High	9.40	9.16	10.88	8.26
Very High	10.46	12.45	8.97	10.12

6. Discussion

It is imperative to map forest fire susceptibility and risk to address the threats they pose to people and the ecosystem. Forest fire dynamics in fire-prone areas can be better understood by identifying potential fire zones. A critical component of forest fire emergency management and mitigation planning is aimed at minimizing adverse effects. However, due to the importance of each conditioning factor in a specific case study area, developing the best FSM methodology presents a challenge. It becomes even more difficult when we have many factors that we are not fully aware of their exact impact on forest fire susceptibility in different regions. In literature studies, the SVM and RF models were found to be appropriate for modeling FSMs. Nevertheless, the accuracy of resulting FSMs will be directly related to the performance of a machine learning model that will, in turn, be affected by the training data set and fire inventory map. It is imperative that every conditional factor must be evaluated so that the training process would be effective. Many conditional factors in forest fire modeling are typically derived from various sources and have been used blindly in many machine learning-based studies. According to previous studies conducted in this specific study area, 17 forest fire conditional factors have been selected for this study. However, in this study, the RFE method was used to select more effective ones. Consequently, the SVM and RF models were first trained with all factors, then with the selected ones.

Seven evaluation criteria were used to validate the effectiveness of the application of this method. Tables 5 and 6 show that the SVM model improved by eight factors in all seven evaluation criteria compared to the SVM model by all factors. Figure eight displays this trend. The improvement rate in the accuracy metric is 10.97%. The AUC metric increased by 8.61%, and other metrics such as recall, precision, F1, Kappa, and MCC improved by 4.98%, 14.17%, 9.87%, 29.13%, and 27.29%, respectively. Our results indicate that the final eight factors have been correctly selected for this model. Also, comparing Tables 3 and 4, the RF model with eight factors involved has improved in all seven evaluation criteria compared to the RF model in which all factors are involved. Accuracy parameter improved by 1.20%, AUC by 1.12%, Recall by 0.69%, Precision by 1.74%, F1 by 1.16%, Kappa by 2.75% and MCC by 2.81%. Figure 9 shows this improvement process. Using this model, the final eight factors have been selected appropriately since they improved all seven factors. Since both SVM and RF models have improved evaluation parameters due to these eight factors, it can be concluded that this feature selection method selects important features of the problem. A feature selection method may be able to improve only one specific model and be unique to that model. But here, from the obtained results, it can be seen that this feature selection method has improved both models, indicating that the problem's important features have been selected by this feature selection method. These features will reduce the accuracy of another model and cannot be generalized to

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

other models if they are added to a second model. However, in this study, the results reveal that the selected factors improve the accuracy of both models, and this feature selection method has worked well in both models. In addition, it is necessary to mention that by comparing Tables 3, 4, 5, and 6 for the RF model in all criteria of evaluation of the model, higher values were found for the RF model in comparison with the SVM model in both cases.

In this regard, the first state of all factors and the second state of eight factors indicate the superiority of the RF model over the SVM model. All these results were achieved on the original dataset and with the 10-fold -cross-validation method. In Table 2, one of the outcomes of this study is identifying the distance from power transmission lines as one of eight factors affecting fire occurrence that is less discussed in other studies. Several studies have identified seven other factors as effective factors in forest fires. (Bjånes et al., 2021; Eskandari et al., 2021; Mohajane et al., 2021; Naderpour et al., 2021; Tavakkoli Piralilou et al., 2022; Valdez et al., 2017). Also, Ghorbanzadeh et al. (Ghorbanzadeh et al., 2019), with 17 factors, achieved the AUCs of 0.79 and 0.88 for the SVM and RF models, respectively. In the current study, the achieved AUCs with eight factors were equal to 0.94 and 0.95 for the SVM and RF models, respectively. This difference in results cannot be categorically stated by choosing this feature selection method because there are differences in the process of the two studies, including the number of folds and also differences in input layers. Furthermore, since the effects of these factors are not completely clear, they cannot be categorically stated what a significant limitation of the problem is.

In general, the identification of high-risk areas is more important than low-risk areas. Therefore, according to Table 7, it can be seen that for the RF model with all factors, the total area of high-risk and very high-risk areas is 19.86%, and for the RF model with eight factors, this number is equal to 19.85%. Also, the SVM model with all factors has identified 21.61% of the region's area as high-risk and very high-risk areas, whereas it was 18.38 for the SVM model with eight factors. In this study, the model that considers all factors was compared to that that only considers a selected number of factors, and the results were very similar for both models. This shows that the selected factors are correctly identified.

7. Conclusions

Our first objective with the present study was to evaluate the selection of the superior forest fire conditional factors for dealing with uncertainties within FSMs using two common ML models, including SVM and RF. Identifying fire-prone areas and preparing a susceptible map for these areas is very important in preventing fires. The prevention and spread of forest fires can be significantly reduced by local and national hazard mitigation management. It is important, however, to identify the practical factors involved in forest fires before identifying areas susceptible to fires. Therefore, this study tried identifying important factors in forest fire occurrence and preparing a fire-prone map for the study area. The influential fire factors were identified by the RFE method based on four models. Then, the most effective factors were selected by voting between factors. The results showed that the accuracy of the two models, SVM and RF, improved. Our applied approach can also be easily adapted to other regions, but factors are needed in each region.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

Author contributions

Ali Rezaei Barzani: methodology, writing – original draft, visualization. Parham Pahlavani: conceptualization, methodology, formal analysis, writing – original draft, supervision. Omid Ghorbanzadeh: conceptualization, methodology, writing – review & editing, visualization, supervision. Pedram Ghamisi: writing – review & editing, supervision, funding acquisition.

Data Availability Statement

The data that support the findings of this study are available on request from the authors.

Acknowledgments

This research was funded by the Institute of Advanced Research in Artificial Intelligence (IARAI). The authors are grateful to the anonymous referees for their valuable comments/suggestions that have helped us improve an earlier version of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

References

Adab, H., Kanniah, K. D., and Solaimani, K.: Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques, Nat Hazards, 65, 1723–1743, https://doi.org/10.1007/s11069-012-0450-8, 2013.

Adab, H., Kanniah, K. D., Solaimani, K., and Sallehuddin, R.: Modelling static fire hazard in a semi-arid region using frequency analysis, Int. J. Wildland Fire, 24, 763, https://doi.org/10.1071/WF13113, 2015.

Agrawal, P., Abutarboush, H. F., Ganesh, T., and Mohamed, A. W.: Metaheuristic Algorithms on Feature Selection: A Survey of One Decade of Research (2009-2019), IEEE Access, 9, 26766–26791, https://doi.org/10.1109/ACCESS.2021.3056407, 2021.

Ampomah, E. K., Qin, Z., and Nyame, G.: Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement, Information, 11, 332, https://doi.org/10.3390/info11060332, 2020.

Bjånes, A., De La Fuente, R., and Mena, P.: A deep learning ensemble model for wildfire susceptibility mapping, Ecological Informatics, 65, 101397, https://doi.org/10.1016/j.ecoinf.2021.101397, 2021.

Bond, W. and Keeley, J.: Fire as a global 'herbivore': the ecology and evolution of flammable ecosystems, Trends in Ecology & Evolution, 20, 387–394, https://doi.org/10.1016/j.tree.2005.04.025, 2005.

Breiman, L.: [No title found], Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Bruinsma, J. (Ed.): World Agriculture: Towards 2015/2030, 0 ed., Routledge, https://doi.org/10.4324/9781315083858, 2017.

Cao, Y., Wang, M., and Liu, K.: Wildfire Susceptibility Assessment in Southern China: A Comparison of Multiple Methods, Int J Disaster Risk Sci, 8, 164–181, https://doi.org/10.1007/s13753-017-0129-6, 2017.

Chicco, D. and Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Genomics, 21, 6, https://doi.org/10.1186/s12864-019-6413-7, 2020.

Devi, S. S., Solanki, V. K., and Laskar, R. H.: Recent advances on big data analysis for malaria prediction and various diagnosis methodologies, in: Handbook of Data Science Approaches for Biomedical Engineering, Elsevier, 153–184, https://doi.org/10.1016/B978-0-12-818318-2.00006-4, 2020.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

Eskandari, S., Pourghasemi, H. R., and Tiefenbacher, J. P.: Fire-susceptibility mapping in the natural areas of Iran using new and ensemble data-mining models, Environ Sci Pollut Res, 28, 47395–47406, https://doi.org/10.1007/s11356-021-13881-y, 2021.

Freund, Y. and Schapire, R. E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, Journal of Computer and System Sciences, 55, 119–139, https://doi.org/10.1006/jcss.1997.1504, 1997.

Geurts, P., Ernst, D., and Wehenkel, L.: Extremely randomized trees, Mach Learn, 63, 3–42, https://doi.org/10.1007/s10994-006-6226-1, 2006.

Ghorbanzadeh, O., Valizadeh Kamran, K., Blaschke, T., Aryal, J., Naboureh, A., Einali, J., and Bian, J.: Spatial Prediction of Wildfire Susceptibility Using Field Survey GPS Data and Machine Learning Approaches, Fire, 2, 43, https://doi.org/10.3390/fire2030043, 2019.

Guo, F., Zhang, L., Jin, S., Tigabu, M., Su, Z., and Wang, W.: Modeling Anthropogenic Fire Occurrence in the Boreal Forest of China Using Logistic Regression and Random Forests, Forests, 7, 250, https://doi.org/10.3390/f7110250, 2016.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V.: [No title found], Machine Learning, 46, 389–422, https://doi.org/10.1023/A:1012487302797, 2002.

He, Q., Jiang, Z., Wang, M., and Liu, K.: Landslide and Wildfire Susceptibility Assessment in Southeast Asia Using Ensemble Machine Learning Methods, Remote Sensing, 13, 1572, https://doi.org/10.3390/rs13081572, 2021.

Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.-X., and Xu, C.: Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China, Science of The Total Environment, 630, 1044–1056, https://doi.org/10.1016/j.scitotenv.2018.02.278, 2018.

Jaafari, A. and Pourghasemi, H. R.: Factors influencing regional-scale wildfire probability in Iran: an application of random forest and support vector machine, in: Spatial modeling in GIS and R for Earth and environmental sciences, Elsevier, 607–619, 2019.

Jaafari, A., Zenner, E. K., and Pham, B. T.: Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers, Ecological Informatics, 43, 200–211, https://doi.org/10.1016/j.ecoinf.2017.12.006, 2018.

Jahdi, R., Salis, M., Darvishsefat, A. A., Alcasena Urdiroz, F. J., Etemad, V., Mostafavi, M. A., Lozano, O. M., and Spano, D.: Calibration of FARSITE fire area simulator in Iranian northern forests, Other Hazards (e.g., Glacial and Snow Hazards, Karst, Wildfires Hazards, and Medical Geo-Hazards), https://doi.org/10.5194/nhessd-2-6201-2014, 2014.

Kalantar, B., Ueda, N., Idrees, M. O., Janizadeh, S., Ahmadi, K., and Shabani, F.: Forest Fire Susceptibility Prediction Based on Machine Learning Models with Resampling Algorithms on Remote Sensing Data, Remote Sensing, 12, 3682, https://doi.org/10.3390/rs12223682, 2020.

Kohavi, J.: Kohavi R., John GH, Wrappers for feature subset selection, Artificial Intelligence, 97, 273–324, 1997.

Kuhn, M. and Johnson, K.: Applied Predictive Modeling, Springer New York, New York, NY, https://doi.org/10.1007/978-1-4614-6849-3, 2013.

Liang, Z., Wang, C., Duan, Z., Liu, H., Liu, X., and Ullah Jan Khan, K.: A Hybrid Model Consisting of Supervised and Unsupervised Learning for Landslide Susceptibility Mapping, Remote Sensing, 13, 1464, https://doi.org/10.3390/rs13081464, 2021.

Mohajane, M., Costache, R., Karimi, F., Bao Pham, Q., Essahlaoui, A., Nguyen, H., Laneve, G., and Oudija, F.: Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area, Ecological Indicators, 129, 107869, https://doi.org/10.1016/j.ecolind.2021.107869, 2021.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

Naderpour, M., Rizeei, H. M., and Ramezani, F.: Forest Fire Risk Prediction: A Spatial Deep Neural Network-Based Framework, Remote Sensing, 13, 2513, https://doi.org/10.3390/rs13132513, 2021.

Ngoc Thach, N., Bao-Toan Ngo, D., Xuan-Canh, P., Hong-Thi, N., Hang Thi, B., Nhat-Duc, H., and Dieu, T. B.: Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study, Ecological Informatics, 46, 74–85, https://doi.org/10.1016/j.ecoinf.2018.05.009, 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, https://doi.org/10.48550/ARXIV.1201.0490, 2012.

Pham, B. T., Jaafari, A., Avand, M., Al-Ansari, N., Dinh Du, T., Yen, H. P. H., Phong, T. V., Nguyen, D. H., Le, H. V., Mafi-Gholami, D., Prakash, I., Thi Thuy, H., and Tuyen, T. T.: Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction, Symmetry, 12, 1022, https://doi.org/10.3390/sym12061022, 2020.

Pourghasemi, H. R., Kariminejad, N., Amiri, M., Edalat, M., Zarafshar, M., Blaschke, T., and Cerda, A.: Assessing and mapping multi-hazard risk susceptibility using a machine learning technique, Sci Rep, 10, 3203, https://doi.org/10.1038/s41598-020-60191-3, 2020.

Pourtaghi, Z. S., Pourghasemi, H. R., Aretano, R., and Semeraro, T.: Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques, Ecological Indicators, 64, 72–84, https://doi.org/10.1016/j.ecolind.2015.12.030, 2016.

Remeseiro, B. and Bolon-Canedo, V.: A review of feature selection methods in medical applications, Computers in Biology and Medicine, 112, 103375, https://doi.org/10.1016/j.compbiomed.2019.103375, 2019.

Sayad, Y. O., Mousannif, H., and Al Moatassime, H.: Predictive modeling of wildfires: A new dataset and machine learning approach, Fire Safety Journal, 104, 130–146, https://doi.org/10.1016/j.firesaf.2019.01.006, 2019.

Song, Y., Niu, R., Xu, S., Ye, R., Peng, L., Guo, T., Li, S., and Chen, T.: Landslide Susceptibility Mapping Based on Weighted Gradient Boosting Decision Tree in Wanzhou Section of the Three Gorges Reservoir Area (China), IJGI, 8, 4, https://doi.org/10.3390/ijgi8010004, 2018.

Tavakkoli Piralilou, S., Einali, G., Ghorbanzadeh, O., Nachappa, T. G., Gholamnia, K., Blaschke, T., and Ghamisi, P.: A Google Earth Engine Approach for Wildfire Susceptibility Prediction Fusion with Remote Sensing Data of Different Spatial Resolutions, Remote Sensing, 14, 672, https://doi.org/10.3390/rs14030672, 2022.

Tehrany, M. S., Jones, S., Shabani, F., Martínez-Álvarez, F., and Tien Bui, D.: A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data, Theor Appl Climatol, 137, 637–653, https://doi.org/10.1007/s00704-018-2628-9, 2019.

Tien Bui, D., Le, K.-T., Nguyen, V., Le, H., and Revhaug, I.: Tropical Forest Fire Susceptibility Mapping at the Cat Ba National Park Area, Hai Phong City, Vietnam, Using GIS-Based Kernel Logistic Regression, Remote Sensing, 8, 347, https://doi.org/10.3390/rs8040347, 2016.

Tonini, M., D'Andrea, M., Biondi, G., Degli Esposti, S., Trucchia, A., and Fiorucci, P.: A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy, Geosciences, 10, 105, https://doi.org/10.3390/geosciences10030105, 2020.

Valdez, M. C., Chang, K.-T., Chen, C.-F., Chiang, S.-H., and Santos, J. L.: Modelling the spatial variability of wildfire susceptibility in Honduras using remote sensing and geographical information systems, Geomatics, Natural Hazards and Risk, 8, 876–892, https://doi.org/10.1080/19475705.2016.1278404, 2017.

Vapnik, V.: The Support Vector Method of Function Estimation, in: Nonlinear Modeling, edited by: Suykens, J. A. K. and Vandewalle, J., Springer US, Boston, MA, 55–85, https://doi.org/10.1007/978-1-4615-5703-6_3, 1998.

https://doi.org/10.5194/egusphere-2022-1294
Preprint. Discussion started: 10 January 2023

Wright, R., Flynn, L., Garbeil, H., Harris, A., and Pilger, E.: Automated volcanic eruption detection using MODIS, Remote Sensing of Environment, 82, 135–155, https://doi.org/10.1016/S0034-4257(02)00030-5, 2002.

Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H., and Hong, H.: Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping, CATENA, 187, 104396, https://doi.org/10.1016/j.catena.2019.104396, 2020.

Yousefi, S., Pourghasemi, H. R., Emami, S. N., Pouyan, S., Eskandari, S., and Tiefenbacher, J. P.: A machine learning framework for multi-hazards modeling and mapping in a mountainous area, Sci Rep, 10, 12144, https://doi.org/10.1038/s41598-020-69233-2, 2020.