Forest wildfire risk mapping, performance comparison of machine learning algorithms

Shadi Jalilian
Gorgan University of Agricultural Sciences and Natural Resources
Shaban Shataee Jouibary (shataee@gau.ac.ir)
Gorgan University of Agricultural Sciences and Natural Resources https://orcid.org/0000-0002-3868-8475

Research Article

Keywords: Artificial neural network (ANN), Fire risk mapping, Multivariate adaptive regression splines (MARS), Random forest (RF), Support vector machines (SVM), Golestan Province

Posted Date: May 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2871614/v1

Abstract

Mapping wildfire risk using proper models and algorithms is one of the top execution priorities for forest managers to prevent wildfires before fires occur. This study evaluates the abilities of the Artificial Neural Network (ANN), Support Vector Machines (SVM), Random Forest (RF), Multivariate Adaptive Regression Splines (MARS) machine learning methods for the prediction and mapping of fire risk across the forests of Golestan Province, Iran. For modeling, the area was first gridded into 1 ha grids, and then pixel values of influential factors were extracted and standardized based on the point shape file of grid centers. The nonparametric algorithms were implemented using 70% of fire points as training samples. The obtained forest fire risk maps were classified into three zones, including low-risk, medium-risk, and high-risk classes. The classification accuracy of the obtained risk maps was evaluated using 30% of the remained fire points. The results showed that the RF algorithm, with an overall accuracy of 75%, had the best performance in fire risk predictions compared to other algorithms. Forest managers can use this methodology to predict areas of most significant fire risk to prevent future fires through land use management, strategic decision-making, and planning. The results enable forest managers to find the best way to monitor, manage, and control fire outbreaks based on fire risk maps of forests in northeastern Iran or other regions with similar conditions.

1 Introduction

Forests are a precious natural resource that plays a crucial part in producing oxygen and purifying the environment (Pourtaghi et al., 2016). Forest biological systems account for around 66% of total earthly carbon and are universally essential to the carbon budget (Ghorbanzadeh et al., 2019). Forests and rangelands have been continuously getting to be imperiled by fires over the world, which have overwhelmingly been activated by anthropogenic variables Cohen et al. (2019). Wildfires are possibly the major destructive natural disaster in forested zones that burn millions of hectares yearly (Boer et al., 2020). Fires destroy biodiversity and imperil the ecosystem slaughtering the considerable mass of inhabitants (Suratman and Abd Latif 2019). Fires have disrupted the natural functioning of local forests. Fire affects not only climatic conditions but also relief formation.

The forests of northeastern Iran, defined as Hyrcanian dense forests, are perfectly adapted to highly harsh climatic conditions. In the Hyrcanian forests and especially in the Golestan Province parts of these forests, wildfires with numerous frequencies have caused substantial losses in forests and other natural resources (Alhaji Khalaf et al., 2021). Mapping the likelihood of fire events occurring and reorganizing high-risk places is fundamental for pre-fire administration and productive firefighting endeavors (Michael et al., 2020). Subsequently, billions of dollars go through yearly on fire management exercises to relieve or avoid wildfires' negative impacts (Jain et al., 2016). To this end, robust approaches and tools are required to enable managers and engineers to estimate the future fire's location, time, and extent (Sevinc et al., 2020). The improvements in techniques for predicting fire susceptibility and delineating the forested zones into different susceptibility levels can help policy and forest managers makers to realize a better understanding of fires that facilitates the improvement of prevention measures for fire-prone forests

(Mhawej et al., 2017). In forest fire prediction, compiling sufficient amounts of data, particularly over large-scale forests, due to field survey difficulties and budgetary constraints (Jaafari et al., 2017).

Over the past decade, machine-learning methods have successfully replaced traditional field survey methods for predicting wildfire exposure by identifying relationships between historical fire events and various explanatory variables for predicting future fires (Pham et al., 2020). Modern tools and methods must be connected to improve wildfire management; in this regard, utilizing artificial intelligence (AI) may be considered an active arrangement (Sayad et al., 2019). AI advances like machine learning (ML), which may be computational considerations of algorithms, can assist researchers in producing strong models for monitoring wildfires and finding (Ghorbanzadeh et al., 2018). The most used ML algorithms were applied in natural hazard and risk modeling and susceptibility mapping studies. Despite the widespread use of these methods, the capabilities of these algorithms are not yet exposed certainly in many parts of the world due to some reasons, such as input data for modeling varies from region to region (Adab et al., 2018; Viedma et al., 2018; Jaafari et al., 2019; Tehrany et al. 2019). To fill this significant gap in fire prediction efforts, we tested a set of predictive ML models based on four machine learning algorithms, including the artificial neural network (ANN), support vector machines (SVM), random forest (RF), and multivariate adaptive regression splines (MARS) for predicting fire susceptibility as the goal of this study. These machine-learning methods have previously been used to model fire risk in other studies (Reyes Bueno and Loján-Córdova, 2022; Eskandari et al., 2020; Milanović et al., 2022; Shao et al., 2022; Xie et al., 2022; Gholamnia et al., 2020; Pang et al., 2022; Akıncı et al., 2023; Rubí et al., 2023).

Although these methods have been extensively explored in environmental studies, particularly for landslide and flood prediction, their combined application and comparison for wildfire risk mapping has not been reported for the entire forests of Golestan Province based on our literature reviews, and most of the current studies have focused on some specific regions. Therefore, the results of previous research are localized and limited, and there is a lack of studies investigating the most suitable and high-precision forest fire prediction model. We selected various forest fire factors that affected it and built four prediction models based on machine learning algorithms and based on what previously occurred fired reported officially for 12 years from 2009 to 2020. The results of this study will allow researchers to determine whether specific predictive models derived from machine learning algorithms are appropriate for their goals of modeling and mapping wildfire risk. Results can also be used to determine the most effective fire risk predictors. These results show the best models for fire risk mapping in this region and similar regions.

2 Materials and Methods

2.1 Study Area

The Golestan Province, located in the northeast of Iran, covers a region of around 2,037,809 ha, in the southeast of the Caspian Sea, comprising 45000 ha of forest and 830000 ha of rangeland. This semi-arid region's thick woodlands and rangelands have long been conducive to wildfire. The territory is known to be one of the foremost wildfire-prone districts of Iran (Golestan Natural Resources Administration 2018). The area of study is situated within the confines of the given coordinates: 36° 30’ – 38° 08” N and 53° 51’ – 56° 22’ E meridians, as they are shown in Fig. 1. The study is focused only on the forest area.

2.2 Data Sources and Description

2.2.1 Occurred wildfires

The coordinates of the recorded wildfire points between 2009 and 2020 were obtained from the conservation unit of the General Office of Natural Resources and Watershed of Golestan Province. These points are then used to create the ground truth map (Fig. 2).

The occurrence of forest fires is influenced by spatial factors that generally have different effects on wildfire occurrence and its frequency. The selection of these factors significantly impacts the effectiveness and excellence of forest fire prediction models. Based on previous research, a large number of analyses, and careful consideration of data accessibility, we finally identified a triad of factors and a total of ten variables. Table 1 provides a comprehensive description of all the variables considered.

Table 1
Description of datasets used in the present study

Factors	Data layer	Source	Data type	Resolution, Units	Purpose
Geographical	Elevation	Digital elevation model	Raster	30 m	Elevation map
	Slope	Digital elevation model	Raster	30 m	Slope map
	Aspect	Digital elevation model	Raster	30 m	Aspect map
	Distance from rivers	Digital elevation model	Polyline	-	Distance from rivers map
Climatic	Average maximum temperature	www.irimo.ir	Point	0.1 ° C	Temperature map
	Average monthly rainfall	www.irimo.ir	Point	1%	Rainfall map
	Average relative humidity	www.irimo.ir	Point	0.1 mm	Relative humidity map
Man-made	Distance from roads	Road network data	Polyline	-	Distance from roads map
	Distance from farmlands	Survey of Iran	Polygon	-	Distance from farmlands map
	Distance from residential areas	Survey of Iran	Point	-	Distance from residential areas map

2.2.3 Identification of spatial factors affecting fire occurrence

The first step in creating a fire risk map is identifying the spatial factors that directly or indirectly affect fire frequency occurring and spreading in any region as sensitive places for fire occurrences. These factors generally include elevation, slope and aspect, average maximum temperature, average monthly rainfall, average relative humidity, distance from roads, distance from farmlands, distance from rivers, and distance from residential areas. These factors were prepared from different sources and ways and mapped in the GIS environment (Fig. 3).

2.4 Extraction and Standardization of Values

For modeling, the area was gridded into decision levels of one hectare. The numerical values of the factors in each grid were extracted based on the point shape file of grids and then were standardized using normalization Eq. 1 (Ma et al., 2020) and rescaled in the range of 0 to 1.

$$\mathrm{Xi}^* = \frac{Xi - Xmin}{\mathrm{Xmax} - \mathrm{Xmin}}$$

1

X_i: the values before and after normalization; Xmax and Xmin: expressed the maximum and minimum values of sample data, respectively.

2.5 Forest wildfire risk mapping

Figure 4 shows the technical flowchart of the study using machine learning (ML) algorithms. Machine learning approaches have significantly contributed to prediction systems, improving performance and efficient results. The persistent progression of ML algorithms over the last few years recognized their suitability for different common risk forecasts with an adequate degree of surpassing conventional approaches. The ML exclusively hinges on the inventory data. For this study, we have used four ML algorithms for wildfire susceptibility mapping, including ANN, SVM, RF, and MARS algorithms. The workflow of the spatial prediction of wildfire susceptibility in the forests of Golestan Province is as follows:

(1) Preparing the conditioning factors based on three main factors, namely physiographic, ecological, and human factors.

(2) Create a wildfire dataset of fire points in the forests of Golestan province in recent years.

(3) Implementing the ANN, SVM, RF, and MARS algorithms for wildfire susceptibility mapping.

(4) Validating and performance evaluation of algorithms.

The methodology and experimental results are summarized in the following sections. Additional explanations and discussions regarding using different ML approaches and training datasets can be found in the Discussion and Conclusions sections.

2.5.1 Artificial Neural Network (ANN)

Artificial neural network impersonates the performance of the human brain through a set of nodes that are interconnected (Ghorbanzadeh et al., 2018). The ANN copies the human brain in two primary regards: firstly, it gets information through a learning strategy; furthermore, the information picked up is put away through synaptic weights (Haykin 2004). The ANN algorithm is prepared to differentiate and disentangle the affiliation between input and yield. In general, there is a connection in multiple-input nonlinear methods among minor single interrelated and interconnected neural systems and weighted interconnects (Fig. 5).

Spatial prediction of wildfire risk could be a complex and nonlinear issue, where an ideal arrangement can be found through ANN by deciding the designs between the conditioning components and reactions (Safi and Bouroumi 2013). For this study, the input layer comprised 55 neurons (based on the number of input conditioning factors), one concealed layer (8 neurons), and one output layer that acted as the network structure. The learning rate was fixed at 0.09, and the number of epochs was set to 500.

2.5.2 Support Vector Machines (SVM)

The SVM is a ubiquitous machine learning approach for data mining, with a set of line indicator features used to evaluate features (Vapnik 2013). The SVM is also known as the maximum margin method that gives better performance and superior results when low data points exist. The SVMs are based on the theory of statistical knowledge, which maps a data set to a multidimensional feature space using a nonlinear transformation to generate the best possible hyperplane (Kavzoglu and Sahin 2013). The SVM has two levels that can implement different functions, such as linear, radial, polynomial, or sigmoid. Therefore, they are one-way. The performance of SVM models is strongly influenced by kernel functions, i.e., linear, polynomial, sigmoid, and radial basis functions (RBF) (Bui et al., 2016). SVMs work as follows (Fig. 6). Classification is performed based on wildfire inventory data, fed into the model as a layer of 1 for wildfire pixel values and 0 for other areas.

2.5.3 Random Forest (RF.)

Random forest (RF) is a machine learning method in which the input dataset is classified based on a gathering of different choice trees. The RF algorithm has gotten expanding consideration for a long time due to its capacity to deliver fantastic classification outcomes with a fast preparation speed (Pourghasemi and Rahmati 2018). Moreover, the highlight set is chosen randomly at each arrangement at which the yield is forecasted, and after that, each yield is given a weighting with esteem based on the votes collected. Based on the yields of choice tree appraisals, the majority vote meets on a single decision tree for the ultimate Classification (Xu et al., 2018). To overcome the uncertainty issue, a single decision tree can be utilized, resulting in higher prediction accuracy (Valdez et al., 2017). RF is considered one of the best nonparametric synchronous learning methods for forest fire susceptibility modeling and mapping purposes. The main training options for RF models are using the maximum number of trees, the number of variables required for split search, and sampling process options (Tavakkoli Piralilou et al., 2019). The maximum number of trees used for this analysis was set to 500. Statistical sampling out of the box (OOB) was used for the final forest model. These sample OOB statistics determine how the model performs given new inputs. The inputs used for training samples are called batch observations, and the inputs are called batch data for the decision tree in the RF approach. Results are the average of all trees results (Fig. 7).

2.5.4 Multivariate Adaptive Regression Splines (MARS)

MARS is a nonparametric and nonlinear machine-learning model (Friedman, 1991). During the processing period, MARS splits the input data set into subgroups and then fits the data to a Spline regression. MARS models are essential in settings where mapping capabilities depend on subgroups of data collected (York et al., 2006). This machine learning approach divides the wildfire impact space into smaller domains, and within each domain, MARS uses a linear classification model to fit a subset of the data (Bui et al., 2019). The divide-and-conquer approach uses MARS modeling, which divides the training dataset into subdomains (Zhang and Goh 2016). MARS can search for plausible scenarios at all levels and provides solutions for different interactions. The general equation for the MARS function (Eq. 2) is (Rezaie-Balf et al., 2019; Bateni et al., 2019):

$$f\left(x\right)=\beta 0+\sum\nolimits_{m=1}^{M}\beta m\lambda m\left(x\right)$$

2

f(x): dependent variable, β0: constant-coefficient; X: predictor variable; λm: basis function.

2.6 Performance evaluation

The performance of algorithms and the obtained fire risk maps were evaluated using the rest of the 30% of historical fire data that was not entered in the modeling process. The accuracy check aim to evaluate the degree to which the map resulting from the Classification matches the reality on the ground. In this study, to evaluate the accuracy of the Classification, the predicted values of the probability of occurrence of the test data were divided into two categories 0 (no fire) and 1 (fire). The accuracy of the Classification was then evaluated by calculating the overall accuracy criterion (Eq. 3).

$$\text{Overallaccuracy}=\frac{Sumofcorrectlyclassifiedpixels}{ThetotalEquationNumberofpixelstoclassify}$$

3

4 Results

4.1 The temporal distribution of fires

This study was performed to determine the best machine-learning algorithm for fire risk mapping in the forests of Golestan Province in northeast Iran. The temporal analysis results of our study show the implications of wildfire in the studied forests in different years. Figure 8 shows the temporal distribution of fires, revealing that most occurred in 2010, 2015, and 2017.

4.2 Fire risk maps

Fire risk maps were created using four machine-learning algorithms. Each of the four selected ML algorithms was applied using the ten wildfire conditioning factors to derive the susceptibility mapping for each algorithm for Golestan Province forests in Iran. The resulting susceptibility map for each ML approach is shown below in Fig. 9. Susceptibility maps indicate wildfire likelihood based on relevant contributing factors for a given area. Each of the four machine learning approaches was used to generate the susceptibility map that then was classified based on natural discontinuity classification methods and classified into one of three classes (high, medium, or low) to serve as a single classification system for comparison purposes. The modeling using each algorithm outputs probability values representing the probability that each pixel will be burned in the future, given a set of contributing variables. These values were used to elaborate wildfire susceptibility maps. The percentage areas of the fire risk classes (high, moderate, low) areas predicted by each of the four algorithms (Fig. 10) and the fire risk areas predicted by each of the machine learning algorithms were determined (Table 2). In addition, the percentages of the different risk levels of the four algorithms were calculated separately. Figure 9 shows that the three risk levels of ANN account for 24.29% (low), 18.33% (medium), and 57.38% (high), respectively, and the three risk levels of SVM account for 34.23% (low), 28.69% (medium), 37.08% (high), and the three risk classes of RF account for 32.29% (low), 40.44% (medium), 27.27% (high), and the three risk classes of MARS account for 52.64% (low), 34.46% (medium), 12.90% (high).

Table 2
Percentage area of fire risk classes produced by machine learning algorithms

Percentage area of fire risk classes	ANN (%)	SVM (%)	RF (%)	MARS (%)
High	57.38	37.08	27.27	12.90
Moderate	18.33	28.69	40.44	34.46
Low	24.29	34.23	32.29	52.64

(ANN), **(b)** (SVM), **(c)** (RF), and **(d)** (MARS).

Validations of the results of the fire risk classifications using overall accuracy (Table 3) indicate that RF, with the highest overall accuracy values, created the most accurate fire risk maps.

Table 3
Accuracy measures of different machine learning algorithms

Model	Accuracy
ANN	0.72
SVM	0.61
RF	0.75
MARS	0.64

5 Discussion

Fire susceptibility mapping is one of the main topics in fire research, and there are different approaches to doing it. Susceptibility mapping is important in reducing the risk of wildfires which pose high risks to humans and ecosystems. It is a key module of emergency management and mitigation planning to reduce adverse impacts (Haas et al. 2013). In the past, only a few technical approaches were used. However, wildfire risk and susceptibility mapping approaches are based on different datasets, metrics, and algorithms (Tehrany et al., 2013; Koetz et al., 2013). This study aimed to predict fire susceptibility using four machine learning methods, namely The artificial neural network (ANN), support vector machines (SVM), random forest (RF), and multivariate adaptive regression splines (MARS). As shown in previous studies (Elith et al., 2006), machine learning approaches outperform statistical approaches. Goldarag et al. (2016) used neural networks and logistic regression to evaluate fire risk, and Ciesielski et al. (2022) used logistic regression to evaluate the contribution of variables to the occurrence of forest fires by excluding weather features as predictors. You et al. integrated a forest resource inventory database based on four aspects: topography, human activities, climate, and forest-related factors. Hence, the results appear to differ based on locale and components. Zhao et al. (2021) selected topography, vegetation, human, and meteorology factors for forest fire assessment using AHP in Laoshan National Forest Park. Nanjing and Goleiji et al. (2017) considered factors such as slope, elevation, and distance from the road and applied an Analytical Network Process (ANP) to assess wildfire risk in cities. These methods involve assigning subjective factors and do not consider factor weights objectively. We selected the following factors for our study: elevation, slope and aspect, average maximum temperature, average monthly rainfall, relative humidity, distance from roads, distance from farmlands, distance from rivers, and distance from residential areas. Li Haiping et al. (2021) developed a logistic risk assessment model for the forest fires in Liangshan County, assuming major risk factors such as topography, vegetation, weather, and population density, and obtained a final AUC of 0.798. As the above model only described a simple linear relationship between the factors, it did not generalize and resulted in a poor accuracy model. The difference between the machine-learning algorithms proposed in this study is that they can solve complex nonlinear relationships between factors quite well. Xie et al. (2022), For wildfire risk assessment based on an integrated machine learning algorithm, Sichuan Province as the specific study area and obtained MCD64A1 burned area product to extract the extent of burned areas in Liangshan Prefecture from 2011 to 2020, and Shao et al. (2022) conduct a study to develop a forest risk map for China. A base map that combines Visible-Infrared Imaging Radiometer Suite data for 17,330 active fires from 2012 to 2019 with topography, meteorology, socio-economic, vegetation, and other factors closely related to wildfire outbreaks for wildfire modeling and the prediction was used. Four machine-learning models for wildfire prediction were compared. The results show clear seasonal and regional differences in wildfire risk in China. Forest fire prediction is an integral part of emergency response management, and mapping fire-prone areas helps to determine the direction of risk management and wildfire risk reduction. However, the accuracy of any fire risk prediction varies depending on the data quality, the method which is used, and additional input parameters. The temporal analysis results of our study show the impacts of wildfire in the study forests in the years 2011 to 2020, and the results show that most fires occurred in 2010, 2015, and 2017. Some studies using machine learning methods (Jaafari et al. 2019; Tien Bui et al. 2019) showed that the SVM model was unsuitable for fire risk mapping in their study area. However, this may be different for other fields and areas. Jaafari and Pourghasemi (2019) and Bui et al. (2017) mentioned that SVM performs well in the spatial modeling of wildfires. However, the results of our research showed that the RF algorithm was the best at predicting fire risk compared to other algorithms, with an overall accuracy of 75%. Different models have advantages and disadvantages, depending on several factors, such as the availability and amount of training data. However, there is no evidence that a particular model is optimal for particular risks, as it depends on the study area and the data available in that particular area (Field et al., 2020). On the other hand, the number of training data points that generally affect the performance of each ML method is still unclear and is a limitation of the current study. Table 5 shows that many researchers have undertaken numerous studies to evaluate the suitability of various methods.

Table 5
Suitability of classifiers in forest fire modeling according to previous studies

Num	Methods Used	Best Method	Study Area	References
1	SVM, GLM, FDA, RF	FDA	Central Koohdasht in Lorestan Province, Iran	(Eskandari et al. 2020)
2	Bayes network, naïve Bayes, multivariate logistic regression, decision tree	Multivariate logistic regression	Pu Mat National Park, Vietnam	(Pham et al., 2020)
3	Frequency ratio−multilayer perceptron, frequency ratio classification, regression tree, frequency ratio support vector machine, frequency ratio−RF.	Frequency ratio−RF	Tanger-Tétouan-Al Hoceima region, northern Morocco	(Mohajane et al, 2021)
4	ANN, SVM	ANN	Guangxi, China	(Li et al., 2020)
5	RF, SVM MLP, GBDT	RF	mainland China	(Shao et al., 2022)
6	SVM, RF, MARS	SVM	Similipal in the northern part of the Eastern Ghat mountain range of India	(Bera et al., 2022)
7	GBM, RF	GBM	The Lower Silesian Voivodeship of Poland	(Milanovi´c et al., 2022)
8	ANN, DR, LARS, MLP, RF, RBF, SOM, SVM, DT,LR	RF	Amol County, Iran	(Gholamnia et al., 2020)
9	BN, NB, DT, MLP	BN	The Pu Mat National Park in the Nghe An Province of Vietnam	(Pham et al., 2020)
10	ANN, SVM, RF	RF	Mazandaran Province, Iran	(Ghorbanzadeh et al., 2019)

In future work, we hope to obtain and use more accurate plant species, flammability, and soil moisture data to model predictive studies. In subsequent research endeavors, including said data in the forest fire prediction model may prove valuable, warranting consideration. The modeling of large-scale forest fires presents a nonlinear and intricate problem, the assessment and prediction of which remains challenging.

The present investigation employed four machine-learning algorithms to develop a forest fire prediction model. In subsequent research endeavors, it may be advantageous to consider alternative machine learning algorithms that might yield promising results. Moreover, the efficacy of these machine learning algorithms in deciphering spatial heterogeneity is comparatively deficient. Future investigations could utilize geographically weighted regression to construct a forest fire prediction model with heightened precision.

6 Conclusions

Natural areas in northeastern Iran are essential for water and soil conservation. However, the lives of rural people in the region are highly dependent on the resources of natural ecosystems. This dependency is reflected in living space, farmland, rural road networks, etc. Wildfires have become a common global risk, devastating forests, burning hectares of habitat, and causing loss of life. Unfortunately, these components of everyday life increase the likelihood of a human-caused fire. Wildfire forecasting is essential to emergency management, and mapping wildfire-prone areas helps mitigate the impact of wildfires. However, each susceptibility map may differ depending on the input parameters and the methodology used to generate it and may have different accuracy. This study used four machine-learning approaches developed and trained based on the historic wildfire events from 2009 to 2020 and ten relevant conditioning factors for the study area. However, this may differ in other research fields where these ML algorithms may have higher accuracies depending on available conditional factors and training datasets. We built four forest fire prediction models using the following machine learning algorithms: ANN, SVM, RF, and MARS. The evaluation results showed that all models' accuracy was higher than 60%. Thus, these models can be used to build forest fire prediction models. Among the four models, the RF model had the highest comprehensive predictive ability, with an accuracy of 75%. It was, therefore, the optimal choice for a forest fire prediction model in the forests of Golestan province in Iran. We used the RF model to predict the probabilities of forest fires in the forests of Golestan province. We had enough wildfire inventory data to train our machine-learning approach and test our machine-learning model.

In addition, In future research, we would like to consider the entire state in the study area, as risk maps can indicate the location of elements at risk and can be included in generating wildfire risk maps that include risk analysis. Obtaining information about communities in the study area with limited capacity and wildfire prevention capacity is critical for mapping the wildfire risk and will help in planning. We must also consider the seasonal aspects, such as seasonal climate data, to map wildfire risk in future work. Since we used the most crucial condition factors related to wildfire occurrence and more general machine learning approaches, the completed workflow easily applies to fire prevention.

Declarations

Funding

-

Competing Interests

-Author 1 and 2 declare they have no financial interests

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Shadi Jalilian and Shaban Shataee. The first draft of the manuscript was written by Shadi Jalilian and Shaban Shataee commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. Adab, H., A. Atabati, S. Oliveira, and A.M. Gheshlagh. 2018. Assessing fire hazard potential and its main drivers in Mazandaran province, Iran: A data-driven approach. *Environmental monitoring and assessment* 190: 1-20.
2. Akıncı, H. A., and H. Akıncı. 2023. Machine learning-based forest fire susceptibility assessment of Manavgat district (Antalya). *Turkey. Earth Science Informatics* 16(1): 397-414.
3. Alhaji Khalaf, M., Sh. Shataee, R. Jahdi, and V. Bacciu. 2021. Improvement of wildfire spread mapping using the development of custom fire fuel models in the replanted forests. *Forest Systems* 30(2): e008,1-14.
4. Arif, M., K. K. Alghamdi, S. A. Sahel, S. O. Alosaimi, M. E. Alsahaft, M. A. Alharthi, and M. Arif. 2021. Role of machine learning algorithms in forest fire management: A literature review. *J Robotics Autom* 5(1):212-226.
5. Bateni, SM., H.R.Vosoughifar, B. Truce, and D.S. Jeng. 2019. Estimation of clear-water is local scour at pile groups using genetic expression programming and multivariate adaptive regression splines. *Journal of Waterway, Port, Coastal, and Ocean Engineering* 145(1):04018029.
6. Bera, B., P. K. Shit, N. Sengupta, S. Saha, and S. Bhattacharjee. 2022. Forest fire susceptibility prediction using machine learning models with resampling algorithms, Northern part of Eastern Ghat Mountain range (India). *Geocarto International* 1-26.
7. Boer, M.M., V.R. de Dios, and R.A. Bradstock 2020. Unprecedented burn area of Australian mega forest fires. Nat. *Clim. Chang* 10: 171–172.
8. Bui, D.T., Q.T. Bui, Q.P. Nguyen, B. Pradhan, H. Nampak, and P.T. Trinh. 2017. Hybrid artificial intelligence approach using a GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. *Agricultural and forest meteorology* 233: 32–44.
9. Bui, D.T., T.A. Tuan, H. Klempe, B. Pradhan, and I. Revhaug. 2016. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 13: 361–378.

10. Bui, DT., ND. Hoang, and P. Samui. 2019. Spatial pattern analysis and prediction of forest fire using new machine learning approach of multivariate adaptive regression splines and differential flower pollination optimization: a case study at Lao Cai province (Viet Nam). *Journal of environmental management*. 237: 476–487.
11. Ciesielski, M., R. Bałazy, B. Borkowski, W. Szczesny, M. Zasada, J. Kaczmarowski, M. Kwiatkowski, R. Szczygieł, and S. Milanovi´c. 2022. Contribution of Anthropogenic, Vegetation, and Topographic Features to Forest Fire Occurrence in Poland. *iForest-Biogeosciences and Forestry* 15(4): 307–314.
12. Cohen, J., D. Coumou, J. Hwang, L. Mackey, P. Orenstein, S. Totz, and E. Tziperman. 2019. S2S reboot: An argument for greater inclusion of machine learning in sub-seasonal to seasonal forecasts. Wiley Interdisciplinary Reviews: *Climate Change* 10(2): e00567.
13. Elith, J., C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, and A. Lehmann. 2006. Novel methods improve the prediction of species distributions from occurrence data. *Ecography* 29: 129–151.
14. Eskandari, S., H. R. Pourghasemi, and J. P. Tiefenbacher. 2020. Relations of land cover, topography, and climate to fire occurrence in natural regions of Iran: Applying new data mining techniques for modeling and mapping fire danger. *Forest Ecology and Management* 473: 118338.
15. Field, RD 2020. Evaluation of Global Fire Weather Database reanalysis and short-term forecast products. *Nat. Hazard Earth Syst* 20: 1123–1147.
16. Friedmann, J.H. 1991. Multivariate adaptive regression splines. *Ann Stat* 19:1–141.
17. Gholamnia, K., T. Gudiyangada, O. Ghorbanzadeh, and T. Blaschke. 2020. Comparisons of diverse machine learning approaches for wildfire susceptibility mapping. *Symmetry* 12(4): 604.
18. Ghorbanzadeh, O., B. Feizizadeh, and T. Blaschke. 2018. Multicriteria risk evaluation by integrating an analytical network process approach into GIS-based sensitivity and uncertainty analyses. *Geomat. Nat. Hazards Risk* 9: 127–151.
19. Ghorbanzadeh, O., K. Valizadeh Kamran, T. Blaschke, J. Aryal, A. Naboureh, J. Einali, and J. Bian. 2019. Spatial prediction of wildfire susceptibility using field survey GPS data and machine learning approaches. *Fire* 2(3): 43.
20. Ghorbanzadeh, O., T. Blaschke, J. Aryal, and K. Gholaminia. 2018. A new GIS-based technique using an adaptive neuro-fuzzy inference system for land subsidence susceptibility mapping. *Journal of SpatialScience* 63: 1–17.
21. Goldarag, Y.J., A. Mohammadzadeh, and A.S. Ardakani. 2016. Fire risk assessment using neural network and logistic regression. *Journal of the Indian Society of Remote Sensing* 44: 885–894.
22. Goleiji, E., S.M. Hosseini, N. Khorasani, and S.M. Monavari. 2017. Forest Fire Risk Assessment-an Integrated Approach Based on Multicriteria Evaluation. *Environ. Monit. Assess* 189, 612.
23. Golestan Natural Resources Administration (GNRA). 2018. Statistics and Data of Fire in Golestan Province. Gorgan: Protection Unit of GNRA Press.
24. Haas, J.R., D.E. Calkin, and M.P. Thompson. 2013. A national approach for integrating wildfire simulation modeling into Wildland Urban Interface risk assessments within the United States.

Landsc. Urban Plan 119: 44–53.

25. Haykin, S. 2004. Neural Network—A comprehensive foundation. Neural Network. 2: 41.
26. Jaafari, A., and Pourghasemi. HR 2019. Factors influencing regional-scale wildfire probability in Iran: An application of random forest and support vector machine. *Spatial modeling in GIS and R for Earth and environmental* 607–619.
27. Jaafari, A., E.K. Zenner, M. Panahi, and H. Shahabi. 2019. Hybrid artificial intelligence models based on a neuro-fuzzy system and meta-heuristic optimization algorithms for spatial prediction of wildfire probability. *Agricultural and forest meteorology* 266: 198-207.
28. Jaafari, A., S.V. Razavi Termeh, and D.T. Bui. 2019. Genetic and firefly meta-heuristic algorithms for an optimized neuro-fuzzy prediction modeling of wildfire probability. *Journal of environmental management* 243: 358–369.
29. Jain, P., S. C. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan. 2020. A review of machine learning applications in wildfire science and management. *Environmental Reviews* 28(4): 478-505.
30. Kavzoglu, T., E.K. Sahin, and I. Colkesen. 2013. Landslide susceptibility mapping using GIS-based multicriteria decision analysis, support vector machines, and logistic regression. *Landslides* 11: 425–439.
31. Koetz, B, F. Morsdorf, S. van der Linden, T. Curt, and B. Allgöwer. 2008. Multi-Source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. *Forest Ecology and Management* 256: 263–327.
32. Li, H., Z. Long, Z. Yang, Z. Xu, and Y. Li. 2021. Analysis of Forest Fire Risk in Sichuan Liangshan Based on Logistic Model. *J. Saf. Environ* 21: 498–505.
33. Li, Y., Z. Feng, S. Chen, Z. Zhao, and F. Wang. 2020. Application of the artificial neural network and support vector machines in forest fire prediction in the Guangxi Autonomous Region, China. *Discrete Dynamics in Nature and Society* 1-14.
34. Ma, W., Z. Feng, Z. Cheng, S. Chen, and F. Wang. 2020. Study on driving factors and distribution pattern of forest fires in Shanxi province. *J. Cent. South Univ. For. Technol* 40: 57–69.
35. Ma, W.; Z. Feng, Z. Cheng, S. Chen, and F. Wang. 2020. Identifying forest fire driving factors and related impacts in China using random forest algorithm. *Forests* 5: 507.
36. Mhawej, M., G. Faour, and J. Adjizian-Gerard. 2017. A novel method to identify likely causes of wildfire. *Clim. Risk Manag* 16: 120–132.
37. Michael, Y., D. Helman, O. Glickman, D. Gabay, S. Brenner, and I. M. Lensky. 2021. Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time series. *Science of the Total Environment* 764: 142844.
38. Milanović, S., J. Kaczmarowski, M. Ciesielski, Z. Trailović, M. Mielcarek, R. Szczygieł, M. Kwiatkowski, R. Bałazy. M. Zasada, and S. D. Milanović. 2022. Modeling and mapping of forest fire occurrence in the Lower Silesian Voivodeship of Poland based on machine learning methods. *Forests* 14(1): 46.

39. Milanović, S., J. Kaczmarowski, M. Ciesielski, Z. Trailović, M. Mielcarek, R. Szczygieł, and S. D. Milanović. 2022. Modeling and mapping of forest fire occurrence in the Lower Silesian Voivodeship of Poland based on Machine Learning methods. *Forests* 14(1): 46.
40. Mohajane, M., R.D. Costache, F. Karimi, Q.B. Pham, A. Essahlaoui, H. Nguyen, G. Laneve, and F. Oudija. 2021. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecological Indicators* 129: 1–17.
41. Pang, Y., Y. Li, Z. Feng, Z. Feng, Z. Zhao, S. Chen, and H. Zhang. 2022. Forest Fire Occurrence Prediction in China Based on Machine Learning Methods. *Remote Sensing* 14(21): 5546.
42. Pham, B. T., A. Jaafari, M. Avand, N. Al-Ansari, T. Dinh Du, H. P. H. Yen,... and T. T. Tuyen. 2020. Performance evaluation of machine learning methods for forest fire modeling and prediction. *Symmetry* 12(6): 1022.
43. Pourghasemi, H.R., and O. Rahmati 2018. Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena* 162: 177–192.
44. Pourtaghi, Z. S., H. R. Pourghasemi, R. Aretano, and T. Semeraro. 2016. Investigation of general indicators influencing forest fire and its susceptibility modeling using different data mining techniques. *Ecological indicators* 64: 72-84.
45. Reyes-Bueno, F., and J. Loján-Córdova. 2022. Assessment of Three Machine Learning Techniques with Open-Access Geographic Data for Forest Fire Susceptibility Monitoring Evidence from Southern Ecuador. *Forests* 13(3): 474.
46. Rezaie-Balf, M., N. Maleki, S. Kim, A. Ashrafian, F. Babaie-Miri, NW. Kim, I-M. Chung, and S. Alaghmand. 2019. Forecasting daily solar radiation using CEEMDAN decomposition-based MARS model trained by the crow search algorithm. *Energies* 12(8):1416.
47. Rubí, J. N., P. H. de Carvalho, and P. R. Gondim. 2023. Application of machine learning models in the behavioral study of forest fires in the Brazilian Federal District region. *Engineering Applications of Artificial Intelligence* 118: 105649.
48. Safi, Y., and Bouroumi. A. 2013. Prediction of forest fires using artificial neural networks. *Applied Mathematical Sciences* 7: 271–286.
49. Sayad, Y.O., H. Mousannif, and H. Al Moatassime. 2019. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire safety journal* 104: 130–146.
50. Sevinc, V., O. Kucuk, and M. A. Goltas. 2020. A Bayesian network model for prediction and analysis of possible forest fire causes. *Forest Ecology and Management* 457: 117723.
51. Shao, Y., Z. Feng, L. Sun, X. Yang, Y. Li, B. Xu, and Y. Chen. 2022. Mapping China's Forest Fire Risks with Machine Learning. *Forests* 13(6): 856.
52. Suratman, M. N., and Z. Abd Latif. 2020. Introductory Chapter: Managing World's Forests for Sustainable Development. *In Forest Degradation Around the World* IntechOpen.
53. Tavakkoli Piralilou, S., H. Shahabi, B. Jarihani, O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S.R. Meena, and J. Aryal. 2019. Landslide Detection Using Multi-Scale Image Segmentation and Different Machine Learning Models in the Higher Himalayas. *Remote Sens* 11: 2575.

54. Tehrany, M.S., S. Jones, F. Shabani, F. Martínez-Álvarez, and D.T. Bui. 2019. A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using logit boost machine learning classifier and multi-source geospatial data. *Theoretical and Applied Climatology* 137: 637–653.
55. Tien Bui, D., ND. Hoang, and P. Samui. 2019. Spatial pattern analysis and prediction of forest fire using new machine learning approach of multivariate adaptive regression splines and differential flower pollination optimization: A case study at Lao Cai province (Viet Nam). *Journal of environmental management* 237: 476–487.
56. Valdez, M.C., K.T. Chang, C.F. Chen, S.H. Chiang, and J.L. Santos. 2017. Modeling the spatial variability of wildfire susceptibility in Honduras using remote sensing and geographical information systems. *Geomat. Nat. Hazards Risk* 8: 876–892.
57. Vapnik, V., 2013. The Nature of Statistical Learning Theory; Springer Science & Business Media: Berlin, Germany.
58. Viedma, O., I. Urbieta, and J. Moreno. 2018. Wildfires and the role of their drivers are changing over time in a large rural area of west-central Spain. *Scientific Reports* 8(1): 1-13.
59. Xie, L., R. Zhang, J. Zhan, S. Li, A. Shama, R. Zhan, ... and R. Wu. 2022. Wildfire risk assessment in Liangshan Prefecture, China, based on an integration machine learning algorithm. *Remote Sensing* 14(18): 4592.
60. Xu, R., H. Lin, Y. Lü, Y. Luo, Y. Ren, and A. Comber. 2018. A modified change vector approach for quantifying land cover change. *Remote Sens* 10, 1578.
61. York, T., L. Eaves, and E. van den Oord. 2006. Multivariate adaptive regression splines: a powerful method for detecting disease-risk relationship differences among subgroups. *Statistics in medicine* 25(8):1355–1367.
62. You, W., L. Lin, L. Wu, Z. Ji, J. Yu, J. Zhu, Y. Fan, and D. He. 2017. Geographical information system-based forest fire risk assessment integrating national forest inventory data and analysis of its spatiotemporal variability. Ecological Indicators 77: 176–184.
63. Zhang, W., and A.C. Goh. 2016. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci Front* 7(1):45–52.
64. Zhao, P., F., Zhang, H., Lin, and S. Xu. 2021. GIS-Based forest fire risk model: A case study in Laoshan National Forest Park, Nanjing. *Remote Sens* 13: 3704.

Figures

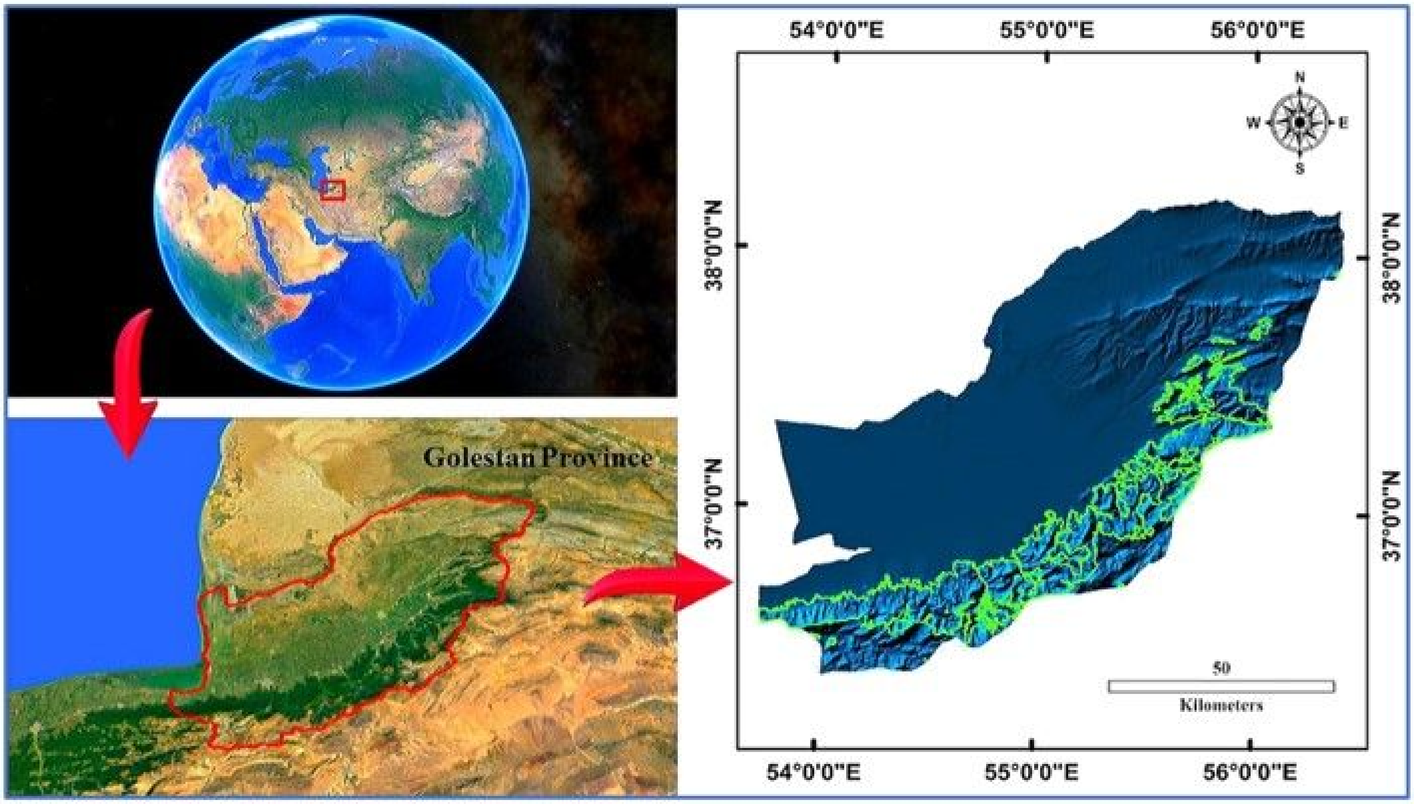


Figure 1

The position of the forests of Golestan province in the study area

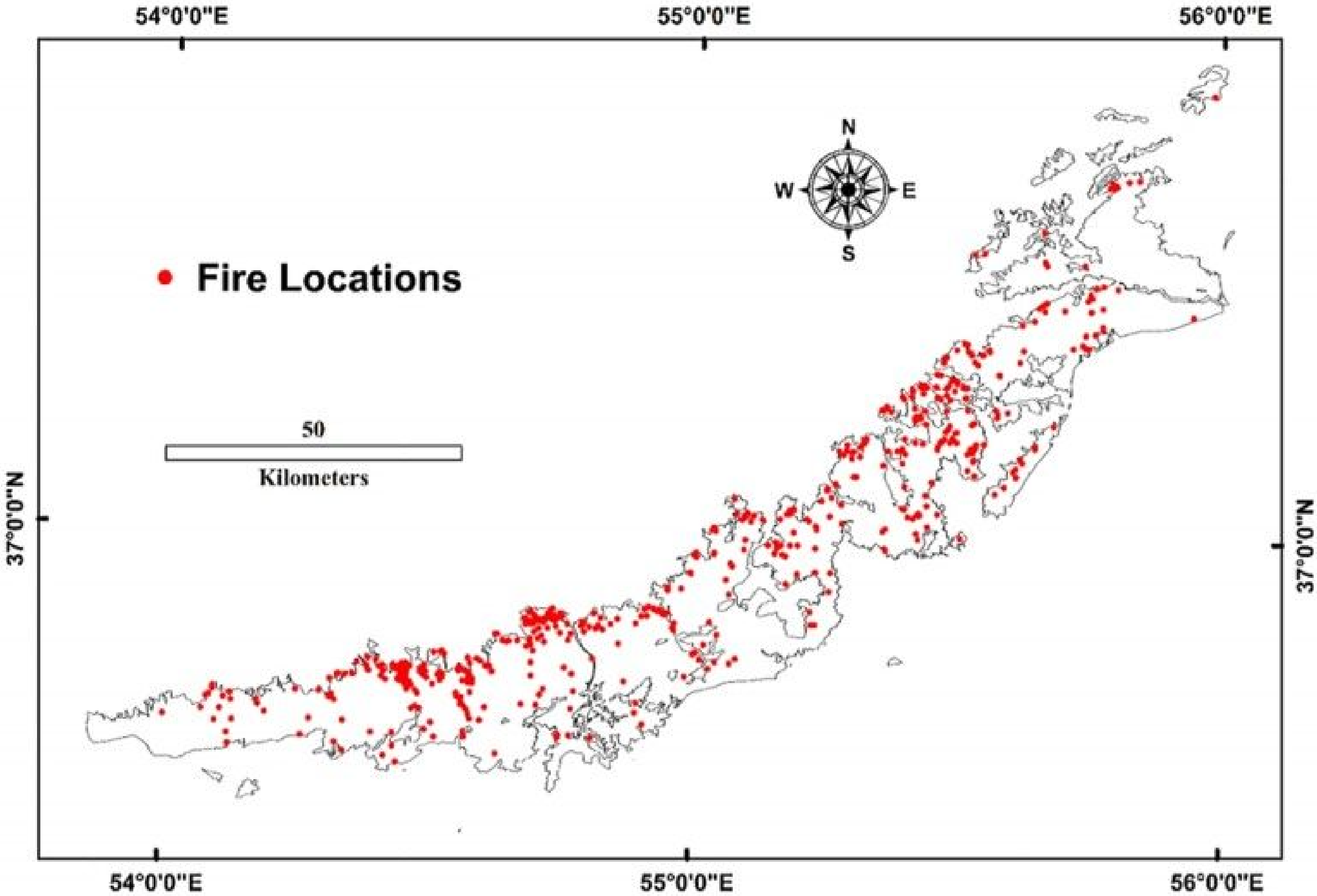


Figure 2

Location of occurred fires between 2009-2020 inthe forests of Golestan Province

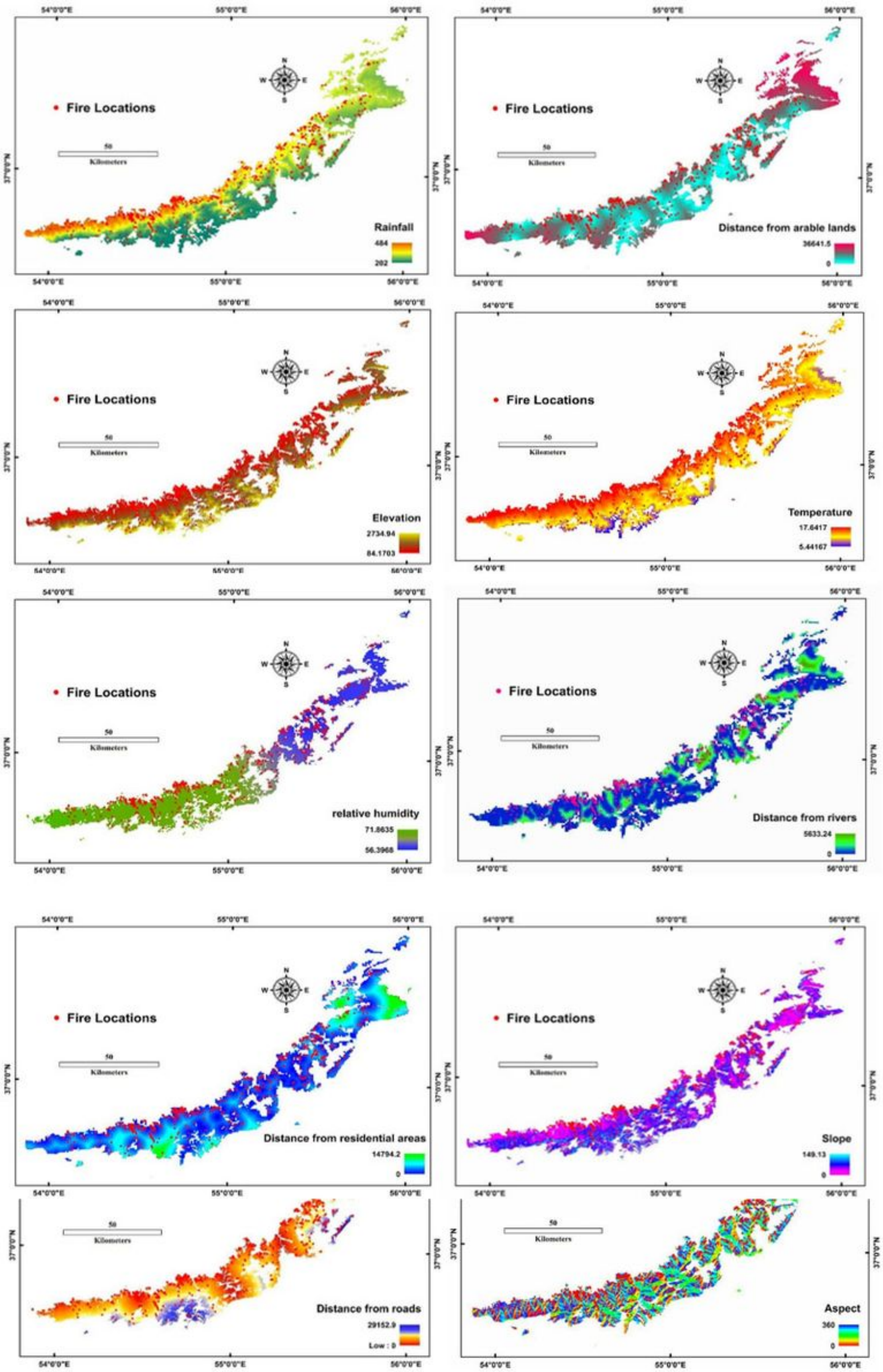


Figure 3

Effective factors on wildfire occurrence

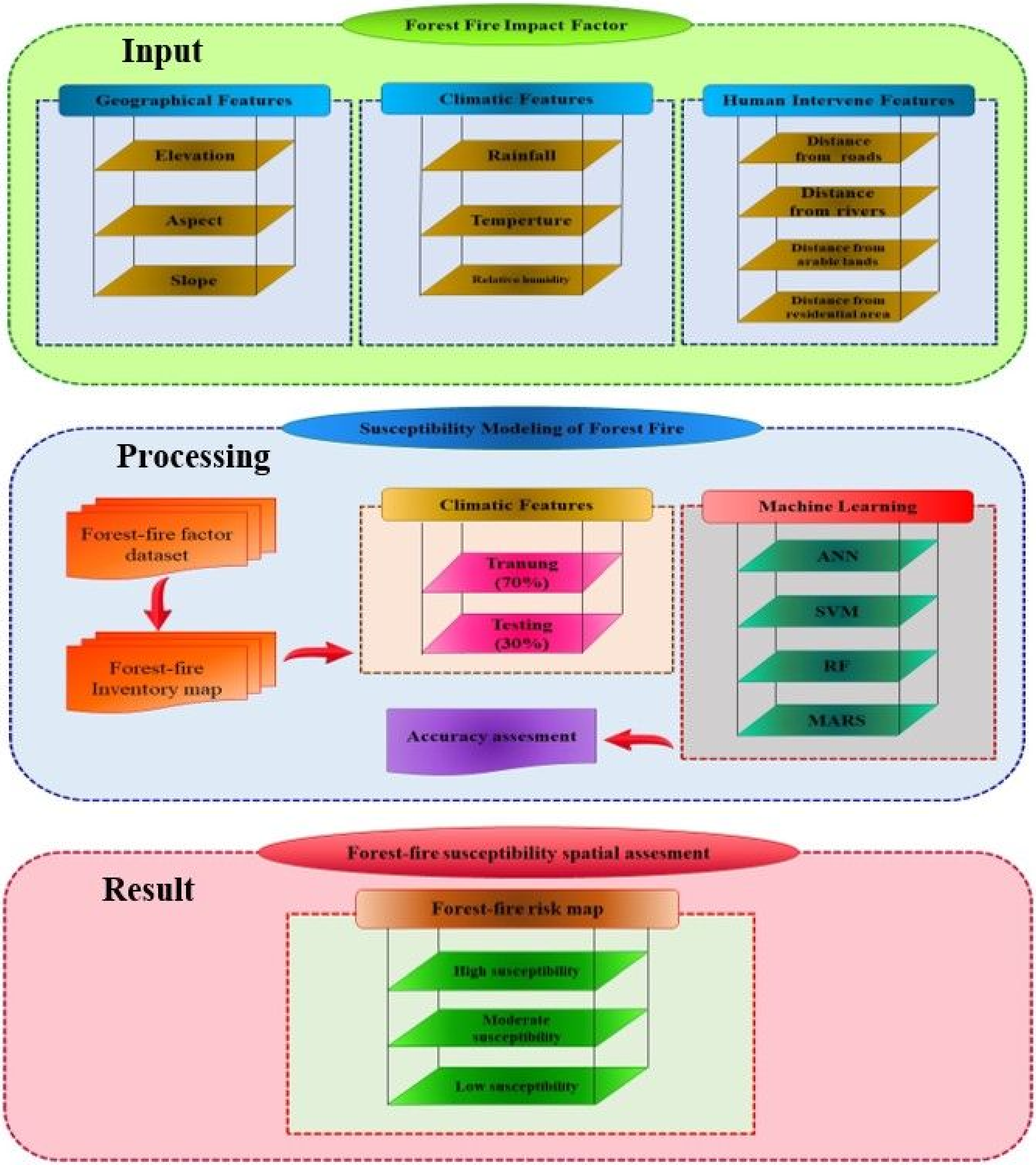


Figure 4

Flowchart of the used analysis

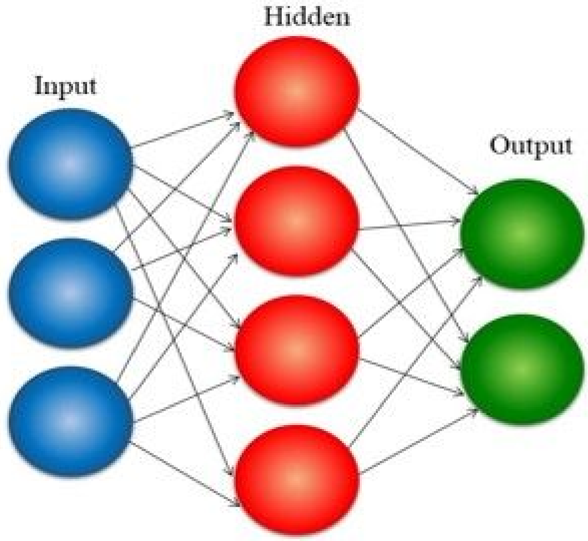


Figure 5

A simple structure of an Artificial Neural Network (ANN) model with input and output layers and only one hidden layer. The input layer consists of the input data (conditional factors), and the output layer is a probability map representing wildfire exposure.

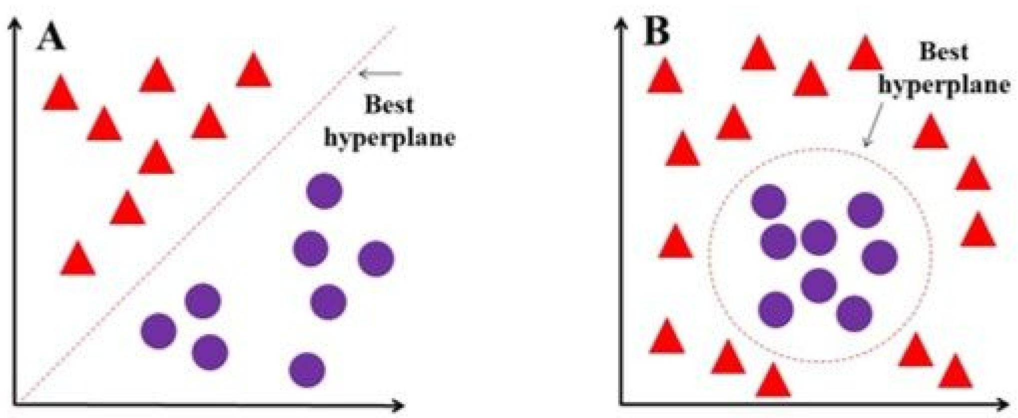


Figure 6

Support Vector Machine (SVM) for continuous**(A)** and Classification **(B)**.

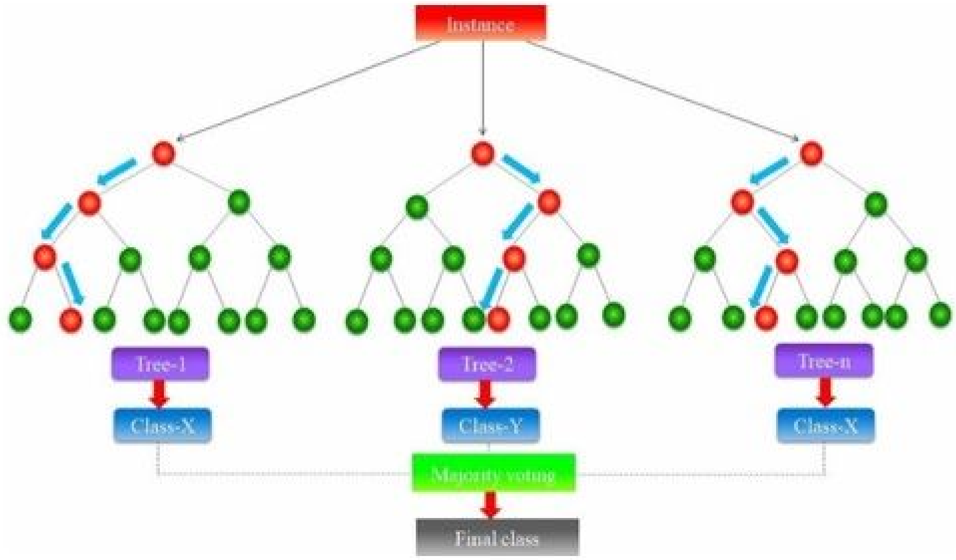


Figure 7

The general structure of the random forest (RF) algorithm

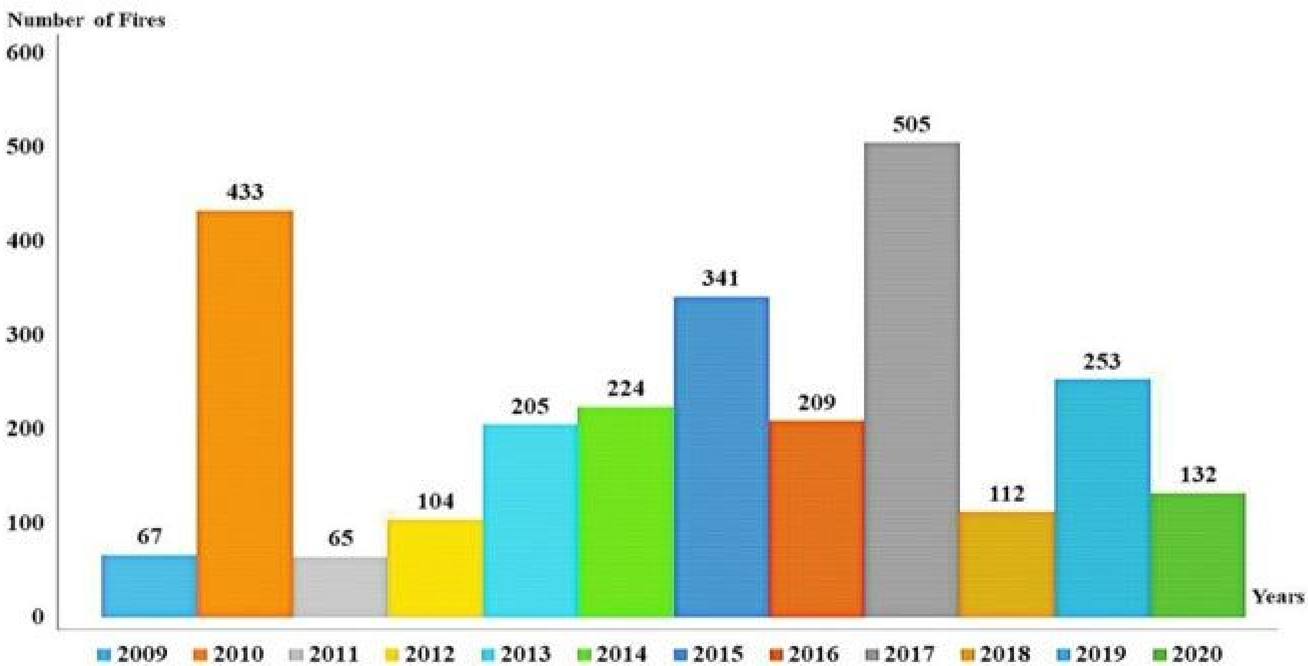


Figure 8

Temporal distribution of fire occurrence in the study area

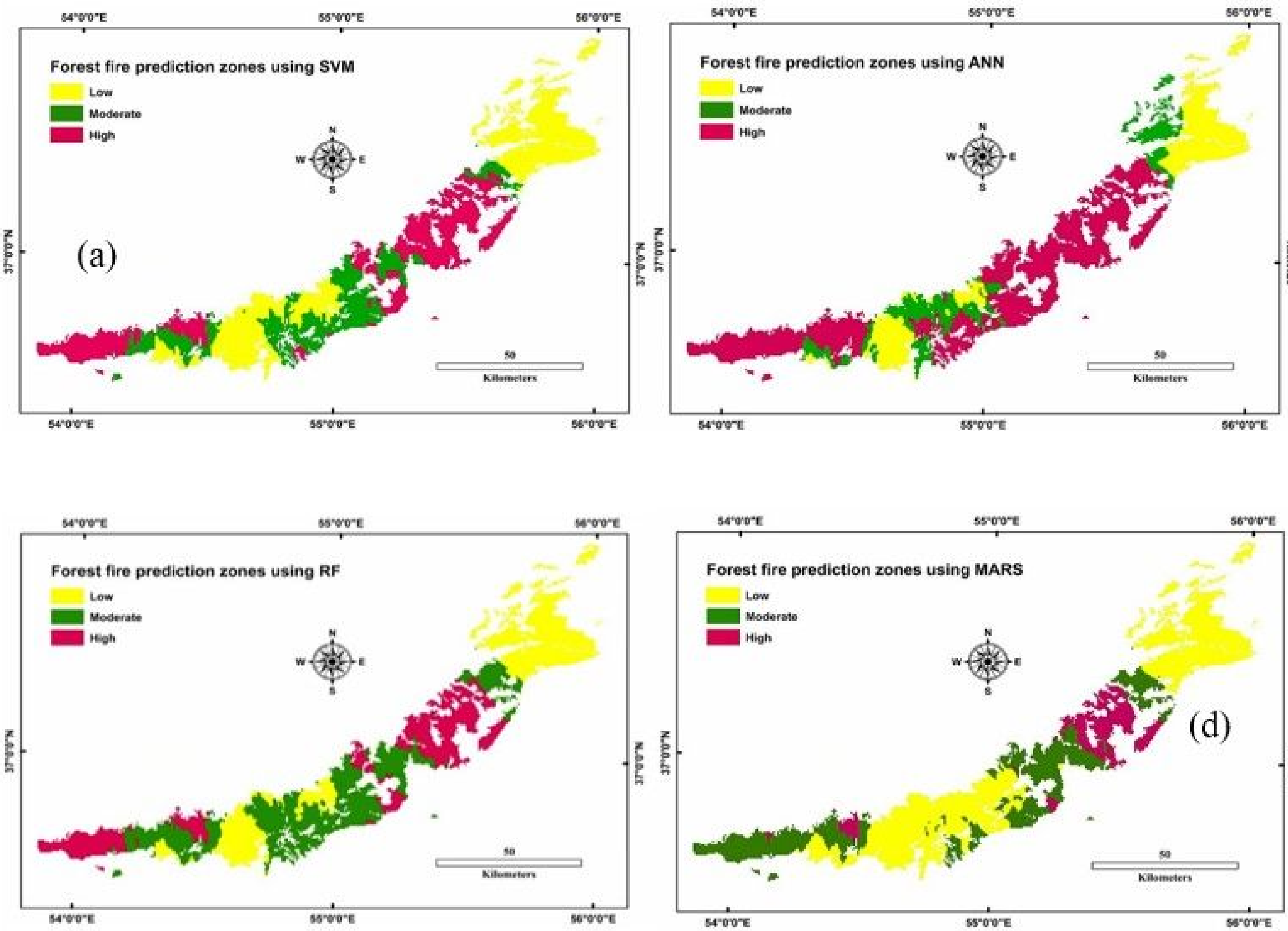
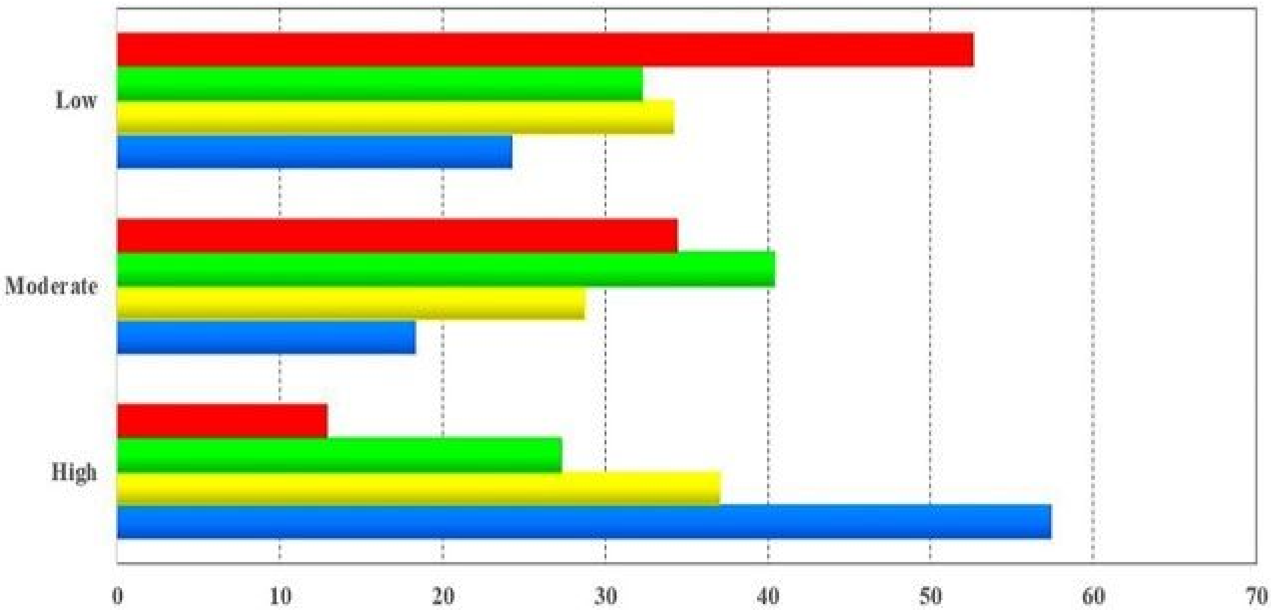


Figure 9

Wildfire susceptibility maps were derived using each machine learning algorithm: **(a)** (ANN), **(b)** (SVM), **(c)** (RF), and **(d)** (MARS).



	High	Moderate	Low
MARS	12.9	34.46	52.64
RF	27.27	40.44	32.29
SVM	37.08	28.69	34.23
ANN	57.38	18.33	24.29

Figure 10

Percentage area of wildfire susceptibility classes by different machine learning algorithms