RESEARCH ARTICLE

Mapping the forest fire risk zones using artificial intelligence with risk factors data

Volkan Sevinç1

Received: 3 December 2021 / Accepted: 9 August 2022 / Published online: 16 August 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Geographical information system data has been used in forest fire risk zone mapping studies commonly. However, forest fires are caused by many factors, which cannot be explained only by geographical and meteorological reasons. Human-induced factors also play an important role in occurrence of forest fires, and these factors depend on various social and economic conditions. This article aims to prepare a fire risk zone map by using a data set consisting of 11 human-induced factors, a natural factor, and temperature, which is one of the risk factors that determine the conditions for the occurrence of forest fires. Moreover, k-means clustering algorithm, which is an artificial intelligence method, was employed in preparation of the fire risk zone map. Turkey was selected as the study area because there are social and economic variations among its regions. Thus, the regional forest directorates in Turkey were separated into four clusters as extreme-risk zone, high-risk zone, moderate-risk zone, and low-risk zone. Also, a map presenting these risk zones were provided. The map reveals that, in general, the western and southwestern coastal areas of Turkey are at high risk of forest fires. On the other hand, the fire risk is relatively low in the northern, central, and eastern areas.

Keywords Forest fire risk zone map · Forest fires · K-means algorithm · Artificial intelligence

Introduction
Forests are important habitats for an ecologically and economically sustainable life. Therefore, it is important to take precautions against various factors damaging the forests. One of the factors that harm the forests is forest fire. Forest fires arise for some natural or unnatural reasons. To prevent forest fires, risk zone maps are useful to take the necessary precautions in accordance with the corresponding risk factors.

In the literature, there are plenty of studies concerning forest fire risk mapping. However, vast majority of these studies use geographical information system (GIS) and remote sensing to prepare these maps.

Among these studies, there are ones using imagery analysis such as Jaiswal et al. (2002), who prepared a forest zones risk map for Gorna Subwatershed in India using GIS data. They digitized the thematic and topographic information and determined the forest fire risk zones by assigning subjective weights to the classes of the layers with respect to their sensitivity to fire or fire-inducing capability. Additionally, Erten et al. (2004) prepared a forest fire risk zone map for Gallipoli area in Turkey by integrating a satellite image, topographical, and other ancillary data obtained from GIS. They prepared the map by performing supervised image classification based on the pixels of the imagery data. Also, Dong et al. (2005) provided a forest fire risk zone map, which was prepared by using ARC/INFO GIS software, for Baihe forestry bureau in China. They interpreted and classified the satellite images to generate vegetation type and land use layers. They prepared the forest fire risk zones by assigning subjective weights to the classes of all the layers regarding their sensitivity to fire or their fire-inducing capability. Afterwards, five categories of forest fire risk ranging from very high to very low were constituted automatically. Nisanci (2010) prepared a fire density map in Trabzon province of Turkey using GIS data and imagery analysis was used to determine the forest fire risk areas. Ghobadi et al. (2012) prepared a forest fire risk zone map for northern forests of Iran using GIS
data. They integrated the parameters affecting the forest fires within GIS using ArcGIS software. Sivrikaya et al. (2014), however, prepared a forest fire risk map of Yeşilova Forestry Enterprise in Kahramanmaraş, Turkey, using GIS data. To create the fire risk map, they began with creating digitized topographic maps of the region. Afterwards, they determined the fire risk index, which is based on vegetation variables, topography, and human factors, by means of a visibility analysis. Bahadır (2010) prepared a forest fire risk zone map of Turkey based on GIS data. The analysis that exists in the study was performed by utilizing GIS techniques. Also, the areal distribution of the forest fires was obtained by a surface analysis and geostatistics was referred to calculate the risks. Aricak et al. (2014) prepared a fire potential map based on stand age, stand closure, and tree species in Kastamonu, Turkey, and they used satellite imagery and performed imagery analysis to prepare the fire potential map. In another study by Gülçin and Deniz (2020), a forest fire risk zone mapping study, which depends on remote sensing and GIS data, was performed in Manisa province of Turkey. They prepared the forest fire risk map using imagery analysis and fire risk zone index calculation using the risk factor scores assigned to the variables. In addition, Sağlam et al. (2008) performed a study to determine spatio-temporal change of fire risk and danger potential in Korudag forest planning unit in Turkey by using Landsat imagery data and remote sensing.

There are also studies using analytic hierarchy process (AHP) techniques such as Sharma et al. (2012), who employed fuzzy AHP techniques for forest fire risk modeling in India. In their study, the forest fire risk zones were designed on knowledge-based information. To improve the efficacy of their model, they applied crisp and fuzzy AHP. Also, Atesoglu (2014) mapped the forest fire risks in Bartın, Turkey, by using GIS data and multi-criteria analysis including AHP. A similar study depending GIS data and AHP was performed by Coban and Erdin (2020) in Bucak, Turkey. Similarly, Pandey and Ghosh (2018) generated a fire risk model to map the fire risks using remote sensing and GIS technique, in Pauri Garwhal District, India. They generated the forest fire risk model by using AHP method, where each category was assigned a subjective weight according to their sensitivity to fire. Additionally, Sivrikaya and Küçük (2022) prepared a forest fire risk map using AHP and statistical analysis in the Mediterranean region. Akbuluk et al. (2018), however, suggested a new approach based on GIS, remote sensing, and AHP to develop a forest fire-risk model. The approach that they propose includes human factors as well as environmental factors.

To analyze fire risks, some studies referred to regression analysis such as Xu et al. (2006), who mapped the forest fire risk zones for Baihe Forestry Bureau in Jilin Province of China using spatial data. They performed a principal component analysis to sort out the relations between forest fire potentials and environmental factors. However, to analyze the relationship between forest fire risks and historical forest fires, they referred to a linear regression analysis. Additionally, Karabulut et al. (2013) determined the forest fire risk zones in Kahramanmaraş province of Turkey by using GIS data. They modeled the forest fire risk factors using an equation employed by Erten et al. (2005) and Joaquim et al. (2007). Mohammadi et al. (2014) also employed a logistic regression model along with GIS for forest fire risk zone modeling in Iran. Bingöl (2017) determined the forest fire risk zones in Burdur province of Turkey with GIS data. To evaluate the risk levels, the equation used by Erten et al. (2005) and Joaquim et al. (2007) was also used in this study. Also, Yathish et al. (2019), built a logistic regression model to make a comparative analysis of forest fire risk zone mapping methods also considering the expert knowledge.

The objective of our study, however, is to prepare a fire risk zone map of Turkey with an artificial intelligence method k-means algorithm, as well as using a data set consisting of variables including human-induced factors, a natural factor, and temperature, which is one of the risk factors that determine the conditions for the occurrence of forest fires (Bilgili et al. 2021).

In this article, the research gap and the motivation of the study are provided in the next section. Afterwards, information about k-means clustering algorithm with its steps and flowchart is provided in the second section. In the third section, the area of study is introduced and the variables used in the study are given with their explanations. Then, in the fourth section, the prepared forest fire risk zone map of Turkey is provided with the colored risk categories and the corresponding zones. Also, various comments about the fire risk zones, some possible reasons for their risk categories, and some differences and similarities among these zones are presented. Finally, some general conclusions are provided in the fifth section.

**Research gap and motivation of the study**

In the literature, the studies concerning forest fire risk zone mapping depend on GIS and remote sensing data. However, forest fires are caused by many factors, which cannot be explained only by geographical and meteorological factors. Human-induced factors, which depend on various social, cultural, and economic conditions, also play an important role in occurrence of forest fires.

A report by the World Health Organization (WHO) suggests that wildfires are often caused by human activity or a natural phenomenon (WHO 2022). Additionally, Food and Agriculture Organization (FAO) reports that worldwide, more than 90% of fires are linked directly or indirectly to intentional and unintentional human actions. Also, the Mediterranean region has the larger proportion of human caused...
fires in the world (95%) followed by South Asia (90%), South America (85%), and Northeast Asia (80%) (FAO 2007). Moreover, according to the US National Park Service (NPS), nearly 85% of wildland fires in the USA are caused by humans (NPS 2022). In a study by Leone et al. (2009), it is stated that socio-economic changes that are occurring in Europe along with global warming result in an augment of fire risk. Additionally, Sevinc et al. (2020) performed a study that estimates the possible causes of forest fires, among which human-induced factors also play a role. Ministry of Agriculture and Forestry (MAF) suggests that forest fires in Turkish forests are mostly caused by humans. In fact, it was recorded that in 2018, 32% of the forest fires were caused by negligence or accident, 19% by natural causes, 4% by intention, and the cause of 45% is unknown (GDF 2019). In another report regarding the socio-economic conditions of the forest villagers in Turkey, which was published by FAO, it is stated that forest villagers in Turkey are poorer than other Turkish villagers. The report also suggests that the number of forest fires increased because of increasing population, industrial development, and mass tourism activities (Kurtulmuslu and Yazici 2003).

The novelty of our article is that it uses a new approach to forest fire risk mapping, as the other forest fire zone mapping studies in the literature mostly depend on GIS or remote sensing data and they use regression models or AHP analysis. However, our study uses k-means clustering algorithm, which is an artificial intelligence method, to prepare a forest fire risk zone map depending on various human-induced and a natural factor, along with temperature variable, which is a risk factor that determines the conditions for the occurrence of forest fires (Bilgili et al. 2021).

Turkey is frequently exposed to forest fires. However, the characteristics of forest fires differ from each other in different regions of Turkey. Moreover, there are also social, cultural, and economic variations among these regions. Therefore, Turkey was chosen as a suitable study area, which allows for different forest fire factors and different levels of damage in different regions to be considered. Thus, in the study, a forest fire risk zone map of Turkey, which depends on 11 human-induced forest fire factors, a natural factor, and temperature as a risk factor determining the conditions for the occurrence of forest fires, was prepared.

Materials and methods

Estimation approaches that are based on artificial intelligence build highly successful results by eliminating some disadvantages of the traditional approaches. Machine learning is a branch of artificial intelligence and there are various machine learning algorithms for different purposes such as classification, clustering, and association. K-means is a machine learning algorithm, which is used for clustering analysis.

In general, there are some advantages of using machine learning methods instead of the classical methods. The most important advantage of machine learning methods, as the name suggests, is their learning capabilities from data without being explicitly programmed. Due to these capabilities, unlike they can increase the accuracy, efficiency, and consistency of the outputs that they produce as new data is added into the model. However, the structures and the mechanisms of the classical models remain the same for every new data set, without getting affected by the past results obtained. Learning and development abilities of machine learning models make them a useful tool for decision making, identification, and tracking without getting affected by human-bias. Therefore, they are also suitable for automation processes, which have a wide range of applications.

K-means clustering algorithm

Clustering analysis is a multivariate statistical method. The purpose of clustering analysis is to bring the variables with similar characteristics in a data set together and collect them in same group called cluster. Thus, various features hidden in large data sets can be revealed with clustering analysis. In clustering analysis, there are two main approaches as hierarchical and non-hierarchical clustering. In the hierarchical clustering analysis, there is no prior information about the number of clusters in the data set and therefore, initially, each observation is assumed to be a cluster. Then, the cluster merge operations are carried out depending on the smallest distance between the cluster centers and this process is continued until the distances are optimally distributed among the clusters (Diday and Simon 1976; Lee 1981; Everitt et al. 2011; Scitovski et al. 2021).

In non-hierarchical clustering analysis, however, the number of clusters k is known or predetermined. K-means clustering is a non-hierarchical clustering algorithm. Each observation in k-means algorithm can only belong to one cluster. K-means algorithm initially determines the randomly selected k cluster center points. At the beginning, the randomly selected cluster center points are randomly selected observations. Afterwards, by assigning the other points to the closest center points, initial clusters are formed. After this process, cluster center points are re-calculated and new clusters are formed by assigning the least distant points to the new clusters. This process continues until the central points of the sets do not change. To calculate the distances, there are some measures such as Euclidian, Manhattan, Canberra, Minkowski, and Mahalanobis distances. Let \( X_i \) and \( Y_i \) denote two data points in a data set of size \( n \). Let \( S \) denote the covariance matrix, \( T \) denote the transpose operator, and \( p \) denote the...
Minkowski distance parameter. Then, Euclidian, Manhattan, Canberra, Minkowski, and Mahalanobis distances can be calculated using the formulas given in Eqs. (1), (2), (3), (4), and (5), respectively (MacQueen 1967; Bock 2008; Jain 2010; Wu 2012; Thakare and Bagal 2015; Blömer et al. 2016; Gupta and Chandra 2020).

\[
\text{Euclidian distance} = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} 
\]

\[
\text{Manhattan distance} = \sum_{i=1}^{n} |X_i - Y_i|
\]

\[
\text{Canberra distance} = \sum_{i=1}^{n} \frac{|X_i - Y_i|}{|X_i| + |Y_i|}
\]

\[
\text{Minkowski distance} = \left( \sum_{i=1}^{n} (|X_i - Y_i|)^p \right)^{1/p}
\]

\[
\text{Mahalanobis distance} = \sqrt{(X_i - Y_i)^T S^{-1} (X_i - Y_i)}
\]

The Minkowski distance is a general form of Manhattan and Euclidian distances. The parameter \( p \) represents the order of the norm and can be any real value. However, \( p = 1 \) is equivalent to the Manhattan distance and the case where \( p = 2 \) is equivalent to the Euclidean distance.

To decide for the number of clusters there are various proposed methods in the literature, such as rule of thumb, information criterion, silhouette method, elbow method, and cross-validation. However, there is not a certain method that guarantees the optimal number of clusters. Still, among these methods, rule of thumb is one of the oldest and it is known to give good results especially for small-sized data sets (Mardia et al. 1979; Ahmed and Mahmood 2015; Hassan et al. 2019). Rule of thumb is applied as follows.

\[
k \approx \sqrt{\frac{N}{2}}
\]

The steps of the k-means clustering algorithm can be summarized as follows.

i. The number of clusters, \( k \), is decided.
ii. \( k \) observations are selected from the data set randomly and these values are assumed to be the initial clusters and their centers.
iii. Each observation is assigned to the closest center point and new clusters are formed.
iv. Cluster centers are re-calculated.
v. If the new cluster centers are the same as the previous ones, the clustering process is stopped; if not, the process is continued from step iii.

The steps of the k-means algorithm given above can also be visualized by a flowchart. The flowchart of the k-means clustering algorithm is provided in Fig. 1.

**Study area and data**

In Turkey, state forests are owned, managed, and operated by General Directorate of Forestry (GDF), which is a government agency. GDF is divided into 30 Regional Forest Directorates (RFD) as of the year 2022. The data used in the study were derived from the forestry statistics collected and published by GDF (2021) for the years 2007–2019. As Hatay and Sinop are new directorates which were established in 2021, the required data for these directorates have not been published by GDF yet. Thus, the data used in the study belong to the 28 regional directorates, which existed in 2019. These directorates are Adana, Amasya, Ankara, Antalya, Artvin, Balıkesir, Bolu, Bursa, Çanakkale, Denizli, Elazığ, Erzurum, Eskişehir, Giresun, Isparta, İstanbul, İzmir, Kahramanmaraş, Kastamonu, Kayseri, Konya, Kütahya, Mersin, Muğla, Sakarya, Şanlıurfa, Trabzon, and Zonguldak directorates.

The variables used in the study consist of the amounts of burned forest areas (ha), which depend on various human-induced factors and a natural factor, as well as temperature as a risk factor determining the conditions for the occurrence of forest fires. The forest fire causes existing in the data set can be listed as temperature, stubble fire, dump fire, hunting, shepherd fire, cigarette fire, picnic fire, terror, arson, expanding, energy lines, traffic accidents, lightning reasons. It is seen that among the possible reasons, all are human-induced factors except for lightning. Although temperature is not a human-induced factor, it cannot be considered as a risk factor causing forest fires directly. It rather can be classified as a risk factor that is effective on the conditions causing forest fires (Bilgili et al. 2021). The variables and their definitions are briefly presented in Table 1.

Figure 2 shows the burned forest areas (ha), which depend on various causes, in the regional directorates of Turkey between 2007 and 2019.

When the patterns in Fig. 2 are examined, in general, the highest amounts of burned areas are seen in Antalya (937.68 ha), Kahramanmaraş (831.38 ha), Balıkesir (816.17 ha), Muğla (629.49 ha), İzmir (564.35 ha), and Çanakkale (472.51 ha) directorates. For all the directorates, the most harmful factors, which cause the highest levels of burned areas, appear to be lightning (3063.27 ha), stubble fire (1264.99 ha), and energy lines (864.78 ha). When the peak points in Fig. 2 are examined, it is seen that energy lines seem to be an extremely harmful factor in Antalya, as it caused a burned forest area of 478.15 ha in this directorate. Additionally, Balıkesir is the most affected directorate by lightning with a burned area of 598.23 ha. Çanakkale, however, is mostly affected by stubble fire factor with 385.48
ha of burned forest area. Moreover, forest fires caused by arson seem to be mostly seen in Kahramanmaras, where they caused 297.39 ha of burned forest area.

**Results and discussion**

To prepare the forest fire risk zone map of Turkey based on k-means clustering algorithm, the number of the clusters \( k \) was determined by using the rule of thumb formula provided in Eq. (6). As there are \( N = 28 \) cases in the data set, \( k \) was approximately calculated as \( k \approx 4 \). Knime (2021) software and its k-means node facility was employed for the clustering process. During the clustering process, the random initialization option was selected and the static random seed value was set to 10. Additionally, the maximum number of iterations was set to 1000 and Euclidian distance function was used. Therefore, the 28 regional directorates were clustered as four risk zones, which were labeled as extreme-risk zone, high-risk zone, moderate-risk zone, and low-risk zone with respect to the 13 variables used in the study. In accordance with the nature of clustering, the regional directorates within the same zones are expected to have similar characteristics with respect to variables, which are various factors causing or preparing ground for forest fires, used in the clustering process. The forest fire risk zones and the corresponding regional directorates in Turkey are provided in Table 2.
The results presented in Table 2 are visualized in Fig. 3. Due to the large differences between the total amounts of burned areas, the figure was created by taking the logarithms of these values.

When the results provided in Table 2, which were visualized in Fig. 3, are examined, it is seen that the total amounts of forest areas belonging to the directorates in the same cluster are close to each other. This result gives an idea that the clustering process has given appropriate and consistent results. To measure the clustering performance, there are various metrics such as Silhouette
coefficient (Rousseeuw 1987), Caliński-Harabasz coefficient (Caliński and Harabasz 1974), and Davies-Bouldin coefficient (Davies and Bouldin 1979). Silhouette coefficient is a metric that is widely used for evaluating clustering performance, which is calculated as follows.

\[ S = \frac{b - a}{\max\{a, b\}} \]  

where \( a \) is the mean intra-cluster distance and \( b \) is the mean inter-cluster distance to the closest cluster. Also, Silhouette coefficient has the following property.

\[ -1 \leq S \leq 1 \]  

Thus, as the value of \( S \) increases, the clustering performance gets better. Mean Silhouette coefficients are the average of the Silhouette coefficients calculated for each directorate within a cluster. Table 3 presents the mean Silhouette coefficients of the four clusters obtained in this study.

When the mean Silhouette coefficients in Table 3 are examined, it can be concluded that all coefficients are over zero except the coefficient of the extreme-risk zone cluster. The low-risk zone cluster has the highest coefficient as
0.78, which indicates that this cluster was formed quite successfully and significantly. The coefficients of the high-risk zone, the moderate-risk zone, and the extreme-risk zone clusters, however, are 0.17, 0.08, and −0.20, respectively, and all of these scores are around zero. Thus, they indicate a moderately satisfactory clustering performance for the related clusters. Overall, it can be concluded that the clustering performance of the k-means algorithm depending on the chosen hyperparameters is significant, thus satisfactory.

Using the risk zones provided in Table 2 and Fig. 3, it is possible to prepare a forest fire risk zone map of Turkey. Figure 4 presents the risk zone map of Turkey regarding the forest fire risk zones and the corresponding directorates.

When the forest fire risk zone map of Turkey presented in Fig. 4 is examined, it is observed that Antalya, Muğla, İzmir, and Çanakkale directorates, which are in the extreme-risk zone, are located on the western and southwestern coasts of Turkey. The common features of these directorates are that besides being tourism regions, they are regions with high population densities. According to Turkish Statistical Institute (TSI), the total population of the directorates in the extreme-risk zone is 8,624,038, which corresponds to approximately 10% of the population of Turkey as of 2021 (TÜİK 2021). Kurtulmuslu and Yazıcı (2003) reports that the number of forest fires increased because of increasing population, industrial development, and mass tourism activities. These results are all in accordance with the extreme-risk zone presented in Fig. 4. Moreover, the directorates in the extreme-risk zone are also places where also agricultural activities are carried out. Therefore, as can be seen in Fig. 4, the high levels of the forest fire damages caused by stubble fire in Çanakkale (385.48 ha), İzmir (167.64 ha), and Muğla (205.70 ha) directorates are not extraordinary. Another factor causing forest fires and most frequently encountered in the extreme-risk zone is lightning. While lightning caused 331.93 ha of burned forest area in Antalya, it damaged 268.08 ha of forest area in Muğla, and 223.34 ha in İzmir directorates. The second important factor that caused the

| Clusters                  | Mean Silhouette coefficients |
|---------------------------|-----------------------------|
| Cluster 1 (extreme-risk zone) | −0.20                      |
| Cluster 2 (high-risk zone)  | 0.17                       |
| Cluster 3 (moderate-risk zone) | 0.08                       |
| Cluster 4 (low-risk zone)   | 0.78                       |

Table 3 The mean Silhouette coefficients of the four clusters

Fig. 4 Forest fire risk zone map of Turkey

Publisher: Springer
largest burned forest area in the directorates belonging to the extreme-risk zone is the energy lines factor, which caused a burned area of 475.18 ha in Antalya, 58.74 ha in Muğla, and 55.01 ha in İzmir directorates. Additionally, Muğla directorate is the second directorate in Turkey most affected by the arson factor, which caused the burning of 36.01 ha of forest area in this directorate.

Although the amounts of burned forest areas in Kahramanmaraş and Balıkesir directorates, which were labeled as high-risk zone, were close to those of the directorates existing in the extreme-risk zone, these directorates were allocated as a separate cluster by the k-means algorithm applied. The reason for this difference is that the main causes of the forest fires and the amounts of burned areas with respect to the causes in these two directorates are different from each other. The most prominent of these differences is that the amounts of areas burned in forest fires, which were caused by picnic fire, are significantly higher in Kahramanmaraş (21.41 ha) and Balıkesir (106.48 ha) directorates than the vast majority of the other forest directorates in Turkey. Another common characteristic of these two directorates existing in the high-risk zone, is that lightning caused significant amounts of burned areas both in Kahramanmaraş (397.22 ha) and Balıkesir (589.23 ha) directorates. The total areas burned by lightning in the extreme-risk zone is 826.48 ha, while the total in the high-risk zone is 986.45 ha. For this reason, it is thought that another effective factor in the separation of extreme-risk region and high-risk region is lightning. A lightning risk map provided by Turkish State Meteorological Service (TSMS) classifies most parts of both Kahramanmaraş and Balıkesir regions in the same group as 10–20 days in terms of the number of lightning days per year (TSMS 2022). Additionally, unlike the directorates in the extreme-risk zone, no forest fires due to hunting fire were observed in the high-risk zone directorates. Another distinguishing indicator between the extreme-risk zone and the high-risk zone is that Kahramanmaraş directorate has an extreme level of arson problem compared to the directorates in the extreme-risk zone.

Mersin, Adana, Şanlurfa, Bursa, Amasya, Trabzon, and Kayseri directorates exist in the moderate-risk category. Among these directorates, Mersin and Adana also exist on the Mediterranean coast. However, they did not fall into the extreme risk category like Muğla or Antalya. This difference can be explained by one of the economic characteristics of Mersin and Adana. There are various studies suggesting that increasing tourism activities increase the risk of forest fires (Buckley 1991; Sun and Walsh 1998; Belsoy et al. 2012; GhulamRabbany et al. 2013). Similarly, coastal regions around the Mediterranean have environmental problems due to increasing tourism development such as pollution and forest fires (Pavón et al. 2003; Kuvan 2005; Pavlek et al. 2017; Curt and Frejaville 2018; Opitz et al. 2020). However, unlike the zones on the west and south-west coasts of Turkey, there are not high densities of touristic settlements and movements in Mersin and Adana directorates. Instead, these cities are more dependent on industry and agriculture rather than tourism, which brought them together in the same cluster. Moreover, the total amounts of forest fires in Mersin (355.80 ha) and Adana (336.05) is less than that of Antalya, which is 937.68. Thus, these two directorates took place in the moderate-risk zone. As for Bursa, this region has a coast to the Sea of Marmara, like Çanakkale, Balıkesir, Istanbul, and Sakarya directorates. However, while Çanakkale and Balıkesir directorates are in the extreme and high-risk zones, respectively; and Istanbul and Sakarya regions are in the low-risk zone, Bursa directorate, however, was clustered into the medium-risk zone. The reason behind this fact can be explained by examining some distinctive properties of Bursa from the other directorates surrounding the Sea of Marmara. For instance, the amount of burned forest areas due to energy lines in Bursa is 1.59 ha, while it is 71.41 ha in Balıkesir and 20 ha in Çanakkale directorates; on the other hand, these scores are 0.62 ha for Istanbul and 1.37 for Sakarya. These results can be interpreted as Bursa suffers from the forest fires induced by energy lines less than Balıkesir and Çanakkale, but more than Istanbul and Sakarya. The same conclusion can also be made for the picnic fire factor. Picnic fire caused 10.77 ha of burned forest area in Bursa, while 106.48 ha in Balıkesir and 46.20 ha in Çanakkale, but only 5.73 ha in Istanbul and 1.47 ha in Sakarya. A notable property of the other members of the moderate-risk zone Şanlurfa, Amasya, Trabzon, and Kayseri have a common point, which is that there is no burned forest area due to traffic accidents being reported in these directorates. However, Şanlurfa directorate has the second highest amount of burned forest area of 34.50 ha due to terror, while there is an amount of 0.10 ha reported in Adana, and there are not any reported fire events due to this factor in the other moderate-risk zone directorates. On the other hand, arson seems to be a common problem of the moderate-risk zone, as the reported burned areas due to arson are more common in this zone. Except for Kayseri and Trabzon directorates having no burned forest areas due to arson, there are recorded fires in Mersin (10.26 ha), Adana (11.41 ha), Şanlurfa (8 ha), Bursa (1.25 ha), and Amasya (9.45 ha) directorates due to this factor. In fact, Adana directorate is the third most damaged directorate from arson in Turkey after Kahramanmaraş and Muğla directorates. Another common feature of the moderate-risk zone is that the damage caused by forest fires induced by shepherd fire is mostly seen in this zone. Shepherd fire caused 3.58 ha of damage in Mersin, 28.90 ha in Adana, 12.70 ha in Şanlurfa, 24.70 ha in Bursa, and 7.14 ha in Amasya directorates.

The average annual temperature of the moderate-risk zone is 14.6 °C, which is less than that of the extreme-risk zone (16.6 °C) and the high-risk zone (15.7 °C). As far as
the directorates in the low-risk zone are considered, the average temperature in this zone is 12.14 °C, which is less than the ones of all zones. Thus, it is possible to interpret that the temperature variable was a determining factor while creating the clusters of the risk zones.

When the map presented in Fig. 4 is examined, it is observed that most of the directorates constituting the low-risk zone are located in the interior, northern, and eastern parts of Turkey. Thus, it can be interpreted that the formation of the low-risk cluster is affected by the excess of annual precipitation in the northern regions having coasts to the Black Sea or the low temperatures in the central and eastern regions due to their geographical heights (on average, 1205 m, and 1829 m, respectively, Elibüyük and Yılmaz 2010). However, although Kayseri directorate is in the central part and Amasya and Trabzon directorates are in the northern part of Turkey, it is observed that they are in the middle-risk zone. This is due to the fact that in addition to temperature variable, which is a risk factor determining the conditions for the occurrence of forest fires, natural and human-induced variables, also played a role in the clustering process. For instance, Kayseri directorate has considerably more amount of burned forest areas due to stubble fire (44.29 ha) and picnic fire (18.33) than all of the directorates located in the central part of Turkey. Similarly, Amasya and Trabzon directorates suffer from forest fires induced by stubble fire (24.90 ha and 72.8 ha, respectively) significantly more than any of the directorates located in the northern part of Turkey.

Compared to the map provided by Bahadır (2010), the forest fire risk map of Turkey, which was prepared using the k-means method, shows that there are some similarities and differences between the two maps. For example, Bahadır (2010) defines some parts of İzmir, Antalya, and Kahramanmaraş directorates very-high-risk zones, and some parts of Balıkesir and Muğla directorates as high-risk zones, which complies with the map provided in this study. Moreover, Bursa, Şanlıurfa, some parts of Mersin, Kayseri, Bursa, and Amasya directorates were labeled as moderate-risk zones by Bahadır (2010), as in our map. However, Bahadır (2010) defines Çanakkale directorate as a moderate-risk zone, which conflicts with our map defining Çanakkale as an extreme-risk zone. In addition, on the map presented by Bahadır (2010), Ankara directorate is presented as a high-risk zone, unlike our map presenting it as a low-risk zone. Still, while Bahadır (2010) used a data set covering the years 1998–2007, we used a data set covering the years 2007–2019 in our study, and it is possible that there have been changes in the number of especially human-induced forest fires and the amount of burned areas in the past 12 years, which led to some changes in the forest fire risk assessment.

We think that the forest fire risk zones map prepared in this article will be useful for making necessary plans and taking precautions to prevent forest fires. Such that, according to the fire risk levels of the relevant regions, it can be ensured that sufficient amounts of personnel and equipment are available for fire-fighting, protection, and security activities in these regions. Moreover, it is also possible to allocate the existing resources to the relevant zones according to their degrees of forest fire risk.

Conclusions

In conclusion, in this study, a forest fire risk zone map for Turkey was prepared. It was seen that the forest fire risk zone maps existing in the literature were all prepared by using GIS data and remote sensing. However, the map provided in this study was prepared differently, by using various forest fire risk variables and an artificial intelligence method, k-means clustering algorithm. The clusters and the burned areas belonging to the directorates presented in Table 2 and Fig. 3 show that the total amounts of forest areas belonging to the directorates in the same cluster are close to each other. Thus, it can be concluded that the k-means algorithm and the selected parameters during clustering provided appropriate and consistent clusters, which formed an accurate fire risk zone map of Turkey. Overall, the most risky zones in Turkey in terms of forest fires appeared to be the ones in the western and south-western regions, while the less risky zones are in the northern and eastern regions. As a continuation of this study, other forest fire risk maps can be prepared for local zones or for other countries by using the same or different risk factors with appropriate artificial intelligence methods.

Author contribution The sole author of this manuscript is solely responsible for all the contributions made in the manuscript.

Availability of data and materials The data used in this study is publicly available in the reference link provided.

Declarations

Ethical approval This article does not contain any studies with human participants or animals performed by the author.

Consent to participate The sole author of the manuscript consents to participate.

Consent to publish The sole author of the manuscript consents to publish.

Conflict of interest The sole author of this manuscript declares that there is no conflict of interest related this study.
References

Ahmed M, Mahmood AN (2015) Novel approach for network traffic pattern analysis using clustering-based collective anomaly detection. Ann Data Sci 2(1):111–130

Akbulak C, Tath H, Aygün G, Sağlam B (2018) Forest fire risk analysis via integration of GIS, RS and AHP: the case of Çanakkale, Turkey. J Human Sci 15(4):2127–2143

Aricak B, Kuçuk O, Enez K (2014) Determining a fire potential map based on stand age, stand closure and tree species, using satellite imagery (Kastamonu central forest directorate sample). Croatian J For Eng: Theory and Application of Forestry Engineering 35(1):101–108

Atesoglu A (2014) Forest fire hazard identifying. Mapping using satellite imagery-geographic information system and analytic hierarchy process: Bartın, Turkey. J Environ Prot Ecol 15(2):715–725

Bahadir M (2010) Türkiye’de (1998-2007) Görülen Orman Yangınlarının Yüzey ve Rakamsal Sorgulama analizi. Nat Sci 5(3):146–162

Belsoy J, Korir J, Yego J (2012) Environmental impacts of tourism in protected areas, J Environ Earth Sci 2(10):64–73

Bilgili E, Kuçük O, Sağlam B, Coşkuner KA (2021) Chapter 1: Forest fires causes, effects, monitoring, precautions and rehabilitation activities. In: Kavzolju T (ed) Mega forest fires: causes, organization and management. Turkish Academy of Sciences, science and thought series No: 33, Ankara, pp 1–23

Bingöl B (2017) Determination of forest fire risk areas in Burdur Province using Geographical Information Systems. Turk J For Sci 1(2):169–182

Blömer J, Lammersen C, Schmidt M, Sohler C (2016) Theoretical analysis of the k-means algorithm—a survey. In: Algorithm Engineering. Springer, Cham, pp 81–116

Bock HH (2008) Origins and extensions of the k-means algorithm in cluster analysis. Electron J Hist Probab Stat 4(2):1–18

Buckley R (1991) Environmental impacts of recreation in parks and reserves. In: Perspectives in Environmental Management. Springer, Berlin, pp 243–258

Calinski T, Harabasz J (1974) The dendrite method for cluster analysis. Commun Stat Theor Methods 3(1):1–27

Caban H, Erdin C (2020) Forest fire risk assessment using GIS and AHP integration in Bucak forest enterprise, Turkey. Appl Ecol Environ Res 18(1)

Curt T, Frejavielle T (2018) Wildfire policy in Mediterranean France: how far is it efficient and sustainable? Risk Anal 38(3):472–488

Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 2:224–227

Diday E, Simon JC (1976) Clustering analysis. In: Digital pattern recognition. Springer, Berlin, pp 47–94

Dong XU, Li-min D, Guo-fan S, Lei T, Hui W (2005) Forest fire risk zone mapping from satellite images and GIS for Baihe Forestry Bureau, Jilin, China. J For Res 16(3):169–174

Elibüyük M, Yılmaz E (2010) Türkiye’nin coğrafi bölge ve bölümlerine göre yükselti basamakları ve eğim grupları. Coğrafi Bilimler Dergisi 21(1):15–24

Erten E, Kurtun M, Musaoglu N (2004) Forest fire risk zone mapping: The case of Manisa, Turkey. Türkiye Ormançılık Dergisi 21(1):15–24

FAO (2007) Fire management global assessment 2006. In: A thematic study prepared in the framework of the Global Forest Resources Assessment 2005. Food and Agriculture Organization of the United Nations, Forestry Paper 151, Rome

GDF (2019) General directorate of forestry, environmental indicators, forest fires (in Turkish). https://cevreselgostergeler.csb.gov.tr/orman-yanginlari-i-85850. Accessed 5 July 2022

GDF (2021) General directorate of forestry, official statistics (in Turkish). https://www.ogm.gov.tr/tr/ormanlarizmesi-ististikler. Accessed 1 Dec 2021

Ghobadi GJ, Gholizadeh B, Dashliburum OM (2012) Forest fire risk zone mapping from geographic information system in Northern Forests of Iran (Case study, Golestan province). Int J Agric Crop Sci 4(12):818–824

GulhamRabbany M, Afrin S, Rahman A, Islam F, Hoque F (2013) Environmental effects of tourism. Am J Environ Energy Power Res 1(7):117–130

Gülçin D, Deniz B (2020) Remote sensing and GIS-based forest fire risk zone mapping: The case of Manisa, Turkey. Türkiye Ormançılık Dergisi 21(1):15–24

Gupta MK, Chandra P (2020) An empirical evaluation of K-means clustering algorithm using different distance/similarity metrics. In: In Proceedings of ICETIT 2019. Springer, Cham, pp 884–892

Hassan AAH, Shah W, Husein AM, Talib MS, Mohammed AAJ, Iskandar M (2019) Clustering approach in wireless sensor networks based on K-means: Limitations and recommendations. Int J Recent Technol Eng 7(6):119–126

Jain AK (2010) Data clustering: 50 years beyond K-means. Pattern Recogn Lett 31(8):651–666

Jaiswal RK, Mukherjee S, Raju KD, Saxena R (2002) Forest fire risk zone mapping from satellite imagery and GIS. Int J Appl Earth Obs Geoinf 4(1):1–10

Joaquim GS, Bahaaeddin A, Josep RC (2007) Remote sensing analysis to detect fire risk locations. GéoCongrès-2007, Québec

Kababulut M, Karacao A, Gurbuz M, Kizilcelma Y (2013) Determination of forest fire risk areas using geographical information systems in Baskonus Mountain (Kahramanmaras). J Int Soc Res 6(24):171–179

Knime (2021) Knime software. https://www.knime.com/. Accessed 1 Dec 2021

Kurtulmuslu M, Yazici E (2003) Management of forest fires through the involvement of local communities in Turkey. In: Ganz D, Moore P and Reeb D (ed) Community based fire management: case studies from China, The Gambia, Honduras, India, the Lao People's Democratic Republic and Turkey. Food and Agriculture Organization of the United Nations Regional Office for Asia and the Pacific Bangkok, Thailand, pp 119–137

Kuvan Y (2005) The use of forests for the purpose of tourism: the case of Belek Tourism Center in Turkey. J Environ Manag 75(3):263–274

Lee RC (1981) Clustering analysis and its applications. In: Advances in information systems science. Springer, Boston, pp 169–292

Leone V, Lovreglio R, Martín MP, Martínez J, Vilar L (2009) Human factors of fire occurrence in the Mediterranean. In: In Earth observation of wildland fires in Mediterranean ecosystems. Springer, Berlin, pp 149–170

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Lecam L and Meyman J (ed) Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, University of California Press, Berkeley and Los Angeles, 1(14), pp 281–297

Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis academic press inc, 15th edn. London Ltd, London, p 518

Mohammadi F, Bavaghar MP, Shabanian N (2014) Forest fire risk zone modeling using logistic regression and GIS: an Iranian case study. Small-scale For 13(1):117–125
Nisanci R (2010) GIS based fire analysis and production of fire-risk maps: The Trabzon experience. Sci Res Essays 5(9):970–977
NPS (2022) National park service, wildfire causes and evaluations. https://www.nps.gov/articles/wildfire-causes-and-evaluation.htm. Accessed 5 July 2022
Opitz T, Bonneu F, Gabriel E (2020) Point-process based Bayesian modeling of space–time structures of forest fire occurrences in Mediterranean France. Spatial Stat 40:100429
Pandey K, Ghosh SK (2018) Modelling of Parameters for Forest Fire Risk Zone Mapping. ISPRS-Int Arch Photogramm Remote Sens Spat Inform Sci 42(5):299–304
Pavlek K, Biščević F, Furčić P, Grgar A, Gugić V, Malešić N et al (2017) Spatial patterns and drivers of fire occurrence in a Mediterranean environment: a case study of southern Croatia. Geografisk Tidsskrift-Danish J Geogr 117(1):22–35
Pavón D, Ventura M, Ribas A, Serra P, Sauri D, Breton F (2003) Land use change and socio-environmental conflict in the Alt Empordà county (Catalonia, Spain). J Arid Environ 54(3):543–552
Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65
Sağlam B, Bilgili E, Durmaz BD, Kadoğulları A, Kıcık Ö (2008) Spatio-temporal analysis of forest fire risk and danger using LANDSAT imagery. Sensors 8(6):3970–3987
Scitovski R, Sabo K, Martínez-Álvarez F, Ungar Š (2021) Cluster Analysis and Applications. Springer, Dordrecht
Sevinc V, Kucuk O, Goltas M (2020) A Bayesian network model for prediction and analysis of possible forest fire causes. For Ecol Manag 457:117723
Sharma LK, Kanga S, Nathawat MS, Sinha S, Pandey PC (2012) Fuzzy AHP for forest fire risk modeling. Disaster Prev Manag 21(2):160–171
Sivrikaya F, Kıcık Ö (2022) Modeling forest fire risk based on GIS-based analytical hierarchy process and statistical analysis in Mediterranean region. Ecol Inform 68:101537
Sivrikaya F, Sağlam B, Akay AE, Bozali N (2014) Evaluation of forest fire risk with GIS. Pol J Environ Stud 23(1):187–194
Sun D, Walsh D (1998) Review of studies on environmental impacts of recreation and tourism in Australia. J Environ Manag 53(4):323–338
Thakare YS, Bagal SB (2015) Performance evaluation of K-means clustering algorithm with various distance metrics. Int J Comput Appl 110(11):12–16
TSMS (2022) Lightning risk map of Turkey (in Turkish). https://www.mgm.gov.tr/kurumsal/haberler.aspx?y=2012&f=yildirim. Accessed 5 July 2022
TÜİK (2021) Address based population registration system results, 2021 (in Turkish) https://data.tuik.gov.tr/Bulten/Index/?p=Adrese-Dayali-Nufus-Kayiti-Sistemi-Sonuclari-2021-45500. Accessed 5 July 2022
WHO (2022) World Health Organization, Wildfires. https://www.who.int/health-topics/wildfires#tab=tab_1. Accessed 5 July 2022
Wu J (2012) Cluster analysis and K-means clustering: an introduction. In: In Advances in K-means Clustering. Springer, Berlin, pp 1–16
Xu D, Shao G, Dai L, Hao Z, Tang L, Wang H (2006) Mapping forest fire risk zones with spatial data and principal component analysis. Sci China Series E: Technol Sci 49(1):140–149
Yathish H, Athira KV, Preethi K, Pruthviraj U, Shetty A (2019) A comparative analysis of forest fire risk zone mapping methods with expert knowledge. J Indian Soc Remote Sens 47(12):2047–2060

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.