REVIEW ARTICLE

Who are the actors and what are the factors that are used in models to map forest fire susceptibility? A systematic review

Santos Daniel Chicas[1] · Jonas Østergaard Nielsen[1]

© The Author(s) 2022

Abstract

In the last decades, natural fire regimes have experienced significant alterations in terms of intensity, frequency and severity in fire prone regions of the world. Modelling forest fire susceptibility has been essential in identifying areas of high risk to minimize threats to natural resources, biodiversity and life. There have been significant improvements in forest fire susceptibility modelling over the past two decades 2001–2021. In this study, we conducted a systematic literature review of literature covering forest fire susceptibility modelling published during this period. The review provides insights on the main themes of forest fire susceptibility modelling research, the main base input factors used in models to map forest fire susceptibility, the main researchers, the areas where this type of research were implemented, technology and models used. It also highlights collaboration opportunities, and regions, such as Central America and Africa, where mapping of forest fire susceptibility is needed. We argue that such knowledge is crucial in order to identify critical factors and opportunities which can aid in improving factor selection and forest fire management.

1 Introduction

Fires are crucial for many ecosystems as they shape the ecology and evolution of species, maintain the integrity and species composition (Syphard et al. 2007), and change community richness and diversity (Manzo-Delgado et al. 2009). In the last decades, natural fire regimes have experienced significant alterations in terms of intensity, frequency and severity (Vilar del Hoyo et al. 2011). These alterations are attributed to climate change associated with drought and high temperature and anthropogenic activities such as changes in land use and demographics (Tuyen et al. 2021; Dang et al. 2021). The changes of fire

✉ Santos Daniel Chicas
   chicassa@hu-berlin.de

   Jonas Østergaard Nielsen
   ostergaj@hu-berlin.de

[1] IRI-THESys and Geography Department, Humboldt-Universität Zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

    ⚘ Springer

regimes are impacting several domains which include the environment, atmosphere, society and economy (Faramarzi et al. 2021; Bisquert et al. 2013; Michael et al. 2021; Hawbaker et al. 2013). Therefore, it is crucial to understand and better predict forest fires in order to reduce these impacts and improve forest fire management (Manzo-Delgado et al. 2005; Jain et al. 2020).

In an effort to understand and predict forest fires researchers have endeavored in integrating climatic, topography, environmental and anthropogenic input factors into models (Vasilakos et al. 2009). This modelling approach has been used in several of the domains of forest fire research which include but not limited to fire risk, probability and susceptibility; fire behavior prediction; fuel characterization; fire weather and fire detection. However, due to the complex interaction of the climatic, topographic, environmental and anthropogenic factors model accuracy varies and no single models is suitable for application to areas with different environmental setting (Sharifi Hashjin et al. 2012; Wang et al., 2020). Therefore, there has been extensive research on the combination of different methods to development more robust and powerful models which has improved the accuracy of forest fire models (Abedi Gheshlaghi et al. 2021; Eskandari et al. 2020; Tien Bui Dieu et al. 2019; Tuyen et al. 2021; Razavi-Termeh et al. 2020; Tehrany et al. 2019). Moreover, in recent years essential literature reviews have been conducted which provide a better understanding of the forest fire modelling approaches, their implementation and challenges in different domains of forest fire research (Naderpour et al. 2019; Jain et al. 2020; Abid 2021). Although, these reviews have provided essential insights on forest fire modelling methods, they have placed very little emphasis on the factors that are used as inputs into the models and the actors (countries and authors) that are conducting this type of research. Better knowledge of the underlying factors is not only critical for fire management purposes (Costafreda-Aumedes et al. 2017) but also for modelling since model performance is heavily depend on the factors that are used (Wang et al., 2020). The identification of actors is also essential to improve research collaboration and implementation. In forest fire research domains, such as forest fire susceptibility modelling, where there is a lack of a comprehensive review on the factors that are used to map forest fire susceptibility researchers often conduct their own literature reviews in an effort to identify and justify the factors and models that are being used (Kim et al. 2019), which may lead to poor model selection and final models containing irrelevant or noise factors. In order to generate better insights of models, actors and identify the base factors that are used to map forest fire susceptibility a comprehensive literature review is necessary. This review will provide researcher with a broader perspective of the actors, models and main base factors that are used to map forest fire susceptibility and in so doing provide critical information which when combined with the researcher's knowledge of the study area can improve model and factor selection to map forest fire susceptibility.

In forest fires research domains one of the two subdomains with the highest application of modelling methods is wildfire susceptibility (Jain et al. 2020). In this paper, forest fire susceptibility, corresponds to the spatial probability of fire occurrence which is generated by building a spatial fire-susceptibility model using remotely sensed fire data with a combination of climatic, topography, environmental and anthropogenic factors (Jain et al. 2020). Determining fire susceptibility is important because identifying areas of highest risk is essential to minimize threats to natural resources, property and life and saves valuable financial resources (Sakellariou et al. 2019). Although, wildfire susceptibility mapping studies are essential and is a domain that has been studied the most using modelling methods a literature review on the input factors and actors that are involved in this type of research is lacking. Therefore, we conducted a systematic literature review from 2001

to 2021 with the objective of identifying the most important base factors that are used in models to map forest fire susceptibility and identify the actors, models and areas where this type of research is being implemented.

2 Materials and methods

A systematic review of forest fire susceptibility literature was conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses methodology (PRISMA) (Liberati et al. 2009). This method consists of four steps identification, screening, eligibility and inclusion (Fig. 1).

Identification: A literature review to identify relevant articles on fire susceptibility mapping published from 2001 to 2021 was conducted in the following databases: Scopus, Web of Science and Ebscohost Greenfile. These databases were selected as they have a large number of articles published on forest fires. These databases were searched using the following criteria:
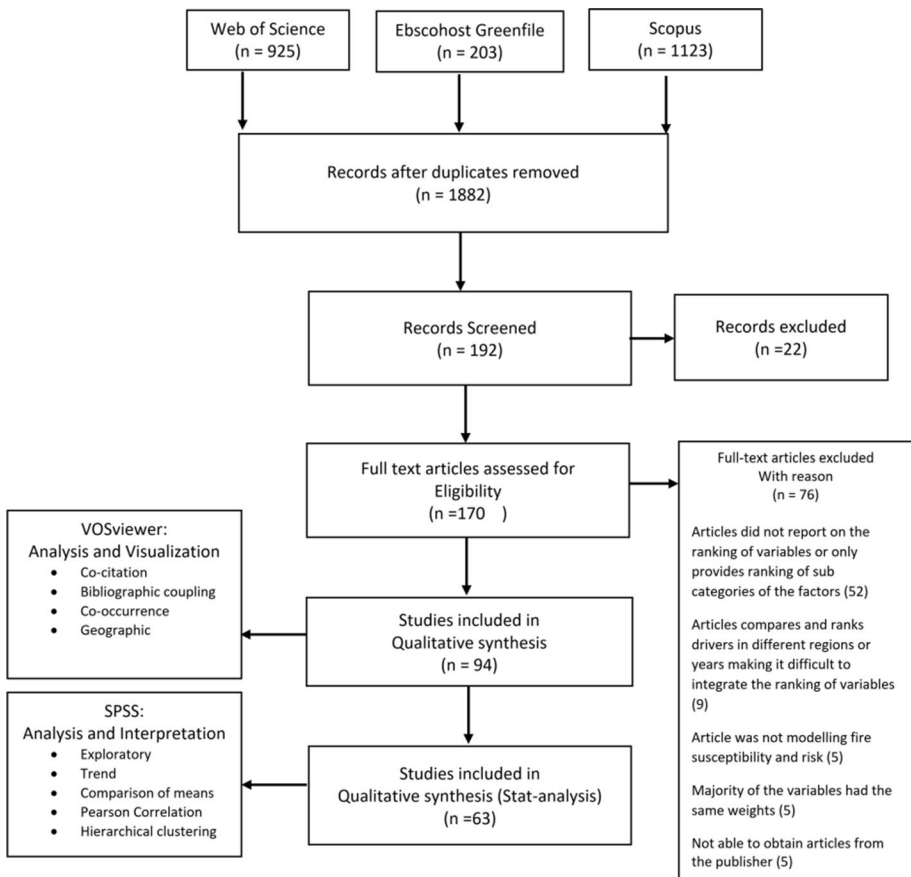

Fig. 1 Methodology

key words "fire" in combination with "wild" or "wildland" or "forest" or "bush" in combination with "susceptibility" or "risk" or "vulnerability" or "mapping" in combination with "parameters" or "factors" or "drivers" or "criteria" or "variables" in combination with "accuracy" or "assessment" or "validation".

This criterion was used searching titles, abstracts, or keywords of the articles in the different databases. There were no language restrictions imposed or paper type and the search was conducted on May, 2021. In total, 2251 articles were identified.

Screening: The title, abstract, keywords, authors' names and affiliations, journal name, and year of publication of the identified records were exported to Endnote and Excel spreadsheet. 369 duplicates were removed. The authors independently screened the remaining 1882 articles by looking at the titles and abstracts of the articles. After this preliminary screening 192 records remained, these were then coded based on relevance following this criterion: 1. Irrelevant, 2. Relevant 3. Unsure. Articles with code one were discarded. Articles categorized with code three were further evaluated by screening the entire content of the articles. At these stage 22 records were excluded.

Eligibility: Since the primary objective of this research was to identify the factors that are used in models to map forest fire susceptibility only articles that comply with the following criteria were considered:

- Utilized a modelling approach (knowledge-based, statistical, machine learning etc.)
- Had information on weight, rank or importance of the factors (topographic, anthropogenic, etc.) used to map forest fire susceptibility. This was to ensure that the factors included in the analysis are those being used in mapping forest fire susceptibility.
- Reported on model accuracy

From the 170 records that were assessed using these criteria. Records that did not contain one of the aforementioned criteria, for the exception of accuracy, were removed. Also, articles that were not in English, reviews or conference articles that were incomplete and articles there we were unable to obtain from the publisher were removed.

Inclusion: After the eligibility assessment, 94 records were included for the bibliographic analysis and of these 63 reported accuracies, which were used for the statistical analysis.

2.1 Bibliographic analysis and visualization: VOSviewer

Bibliometric analysis is a statistical method used to analyze published scientific literature by conducting a statistical evaluation of major trends in publications for a particular topic (Biresselioglu et al. 2020). Clustering techniques are used in bibliometric research because they are useful in identifying the relationship between journals, publications, authors and organizations (van Eck and Waltman 2017). A widely used open source software to conduct bibliometric research is VOSviewer (Briones-Bitar et al. 2020; Tao et al. 2020; Carrión-Mero et al. 2021; Herrera-Franco et al. 2021). This software automatically processes semantic clustering to identify the relationships and relevance between items, thus facilitating the analysis and removing error risk (van Eck and Waltman 2017; Biresselioglu et al. 2020). This study performs a multidimensional bibliometric analysis using VOSviewer software (version 1.6.13). Data extracted from 94 articles included authors, authors affiliation, year of publication, key words and number of citations. From this information,

networks were constructed based on co-citation, bibliographic coupling, and keyword co-occurrence.

2.2 Analysis and interpretation for factors, models and accuracies

Data on area of study, accuracy, modelling method and factors was extracted by way of reading the 94 articles. For the studies that reported more than one modelling method only the information of the model with the highest accuracy was extracted and only factors that were ranked and used by that model were recorded. This ensured that the factors included in the analysis are those being used in modelling forest fire susceptibility. Also, only factors that were used in three or more models across all the reviewed articles were included in the analysis. Factors were grouped by country and a correlation matrix and a hierarchical clustering analysis was conducted to investigate the correlation and relationship of factor in the 94 articles.

Mann Whitney U test, Independent t-test and Kruskal Wallis statistical tests were conducted to determine if model accuracy reported by the 63 articles differs significantly based on study scale, county's level of development and modelling method. Depending on study area scale the study was classified as local, national or regional. National and regional studies were grouped in one category since there were only two regional studies which reported on accuracy. Mann–Whitney U Test was used to compare if there was a significant difference in accuracy as a result of study scale. Countries where the studies were conducted were classified based on the World Banks country classification; thus, countries were placed in three categories: High income countries, upper middle-income countries and lower middle-income countries. There was only one low income country; thus, this county was grouped with the lower middle-income countries. Since the country's level of development had three groups and the data was not normally distributed we conducted the Kruskal Wallis test to determine if there was a significant difference of accuracy as a result of countries level of development (McDonald, 2014). Modelling methods were grouped in two categories: No-machine learning which included: knowledge-based, statistical and others models and Machine Learning which included: machine, hybrid and ensemble models. T test was used since our data was normally distributed to determine if there was a significant difference in accuracy as a result of modelling method used (Poncet et al. 2016).

3 Results

The results of this study are composed of two components. The first component, *Actors (countries and authors)*, consists of a bibliographic analysis of the selected articles. Co-citation analysis, bibliographic coupling, key word analysis and geographic analysis were done. In the second component, *Factors and Models*, we conducted a descriptive and statistical analysis of the factors and models used to map forest fire susceptibility. Combined, we provide insights of the major factors, actors, implementation and collaboration opportunities that can improve forest fire management.
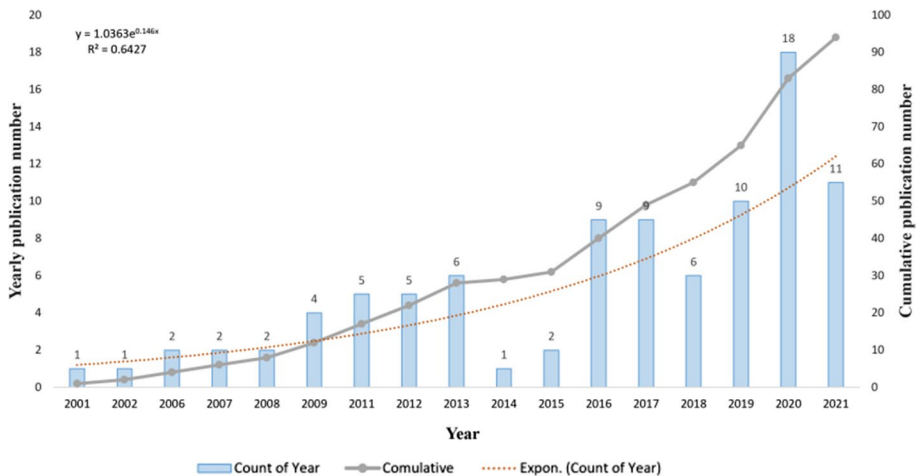
3.1 Publications

During the period 2001–2020 there is an increased interest in using models to map forest fire susceptibility (Fig. 2). For analysis purposes, we divided this period by decades. Period one (2001–2011) and Period two (2012–2021). The average number of publications in period one was two. In comparison to period two, the average number of publications was eight. The total publication contribution for period one and two are 18 and 82%, respectively. The last six years were the most productive years accounting for 67.7% of the publications.

3.2 Actors

3.2.1 Academic productivity, collaboration and geographic analysis

The 10 must productive countries, based on author affiliation, were represented in terms of publications and number of citations. In term of publications, Iran and India are most productive with 28 and 15, respectively (Fig S1b). The US and Iran are the countries with the most citations, 761 and 671, respectively (Fig S1a).

To investigate the connectedness between publications and author's affiliations, a bibliographic coupling network was built. A total of 49 countries were identified; countries with at least two publications were selected to conduct these analyses. Iran, India, US, Spain and Italy are the most productive countries and there are well-established connections between these (Fig. 3). Countries that have recently started scientific production on this topic such as Germany, Austria and Vietnam have strong relationships with the most productive

Fig. 2 Fire susceptibility publications. Price's exponential growth law was used to evaluate scientific production (Montalván-Burbano et al. 2021). Two trend models were developed based on the number of publications, the linear trend model ($y=0.7353x–1.0882$; $R2=0.6408$) and the exponential model ($y=1.0363e0.146x$; $R2=0.6427$). The exponential trend line provided a slightly better fit base, which models the data by 64.27%. From the overall analysis there is an indication that publications in mapping forest fire susceptibility is trending, especially in recent years
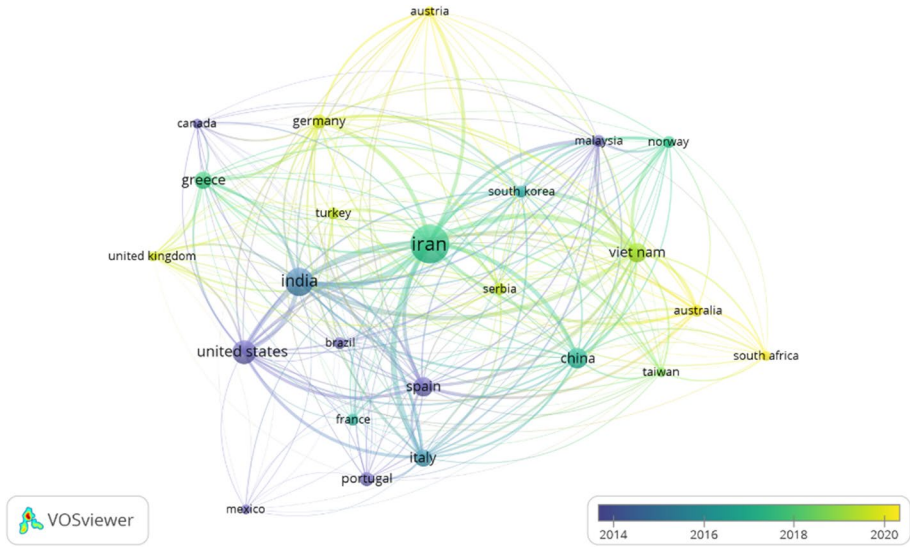
Fig. 3 Network author's affiliation the greater font and circles indicate the most productive countries. The thickness of the line connecting countries indicate the strength of the link between countries considering citations. Countries that started publishing in the earlier years are represented in purple and those who started publishing studies recently on mapping forest fire susceptibility are represented in yellow

counties in the field. Countries such as Taiwan, South Africa and United Kingdom having weak relationships.

Where forest fire susceptibility studies are implemented were extracted from the 94 articles and mapped. Figure 4 shows the 32 countries where forest fire susceptibility research has been conducted. Most of the research has being done in Iran, India and Vietnam with 24, 12 and 7 studies, respectively.

3.2.2 Co-citation and bibliometric coupling analysis

In order to establish the relationship that exists between articles and authors, we conducted a co-citation analysis (Fig. S2). The 94 articles in our database cited 4,565 references, six of the articles met the threshold of four citations for a cited reference. The most connected reference was the article published by (Chuvieco and Congalton 1989) and the second was an article published by (Pourghasemi et al. 2012). For the author analysis, only authors that met the minimum citation of 20 were included, which were 53 authors. The three most important authors were Pourghasemi, Pradhan, and Chuvieco. For the bibliographic coupling article analyses only articles that had a minimum of five citations were included. The 58 articles that met this threshold were used to create a network (Fig S3). The three articles with the highest indices of bibliographic coupling are Pourtaghi et al. (2015), Pourghasemi et al. (2020) and Tehrany et al. (2019). The network also indicates that among the articles the works of Syphard et al. (2007)) and Martínez et al. (2009) are the most cited in the
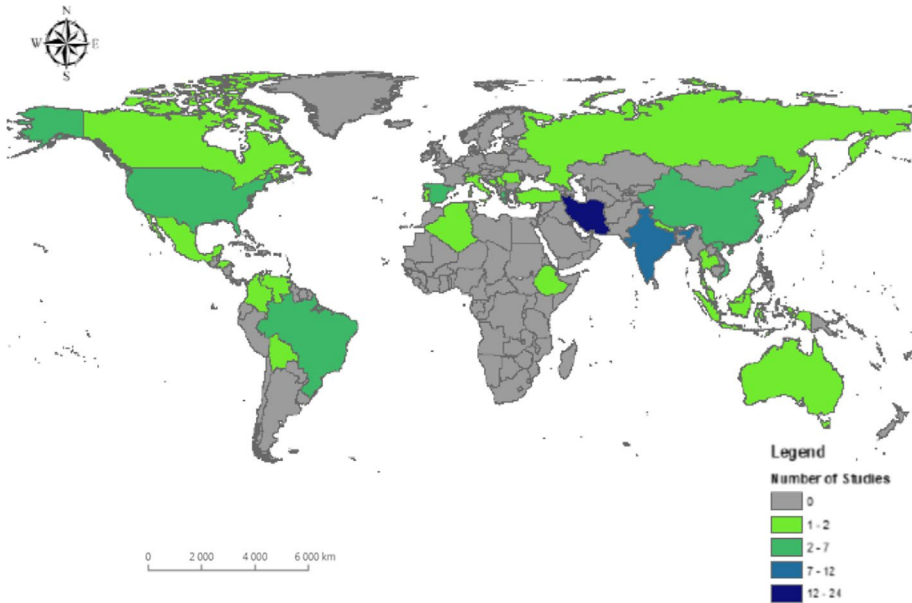
Fig. 4 Location of study implementation

current literature; however, recent work such as Pourtaghi et al. (2016) and Tien Bui D. et al. (2017) are becoming important.

3.2.3 Keyword co-occurrence analysis

We conducted an analysis of keyword co-occurrence between publications. Author keywords and index keywords were extracted from the 94 articles. From the 939 keywords only those that occurred at least five times were kept. The 71 words that met this threshold were further analyzed to remove words that were not relevant (e.g. article, dataset) and duplicates. The 45 remaining keywords were analyzed with VOSViewer's tool for exploring the co-occurrence. The density visualization analysis shows the most frequent
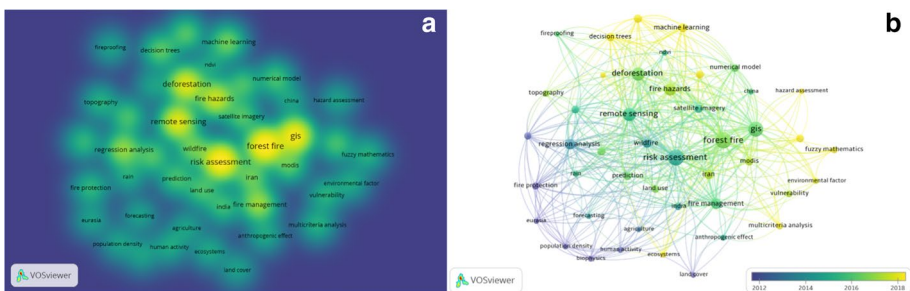
Fig. 5 a Keyword density, **b** Overlay network

keywords: forest fires, risk assessment, GIS, deforestation, remote sensing, fire hazard, regression analysis and fire management (Fig. 5a). The overlay visualization of the keywords in terms of occurrence and average publication year provides a better understanding of the evolution of the methodologies used for mapping forest fire susceptibility. Regression analysis was the method that was predominant in the earlier years. Recently there has been a diversification of methodologies used (Fig. 5b). The most predominantly method implemented recently is machine learning. In the years 2001–2013, the focus is on factors such as population density, human activity and landcover. From 2014 to 2017, the focus is on technology with keywords such as remote sensing, GIS, satellite imagery and MODIS. In 2018–2021, the focus is models with keywords such as machine learning, multicriteria analysis, decision trees and random forest.

3.3 Factors and models

3.3.1 Factors

From the 94 articles, 144 factors used as inputs in models to map forest fire susceptibility were extracted. In order to extract the most important factors used in models to map forest fire susceptibility only those used in three or more models were extracted. In total, 33 factors met this criterion (Table 1). The factors that were used the most were slope (68), elevation (67) and distance to roads (67) and the factors that were used the least were soil moisture (3), shrub (3) and livestock density (3) (Fig S4). The 10 most commonly used factors account for 70% of the factors used as inputs in models to map forest fire susceptibility in the past twenty years. From the 676 times that factors were used topographic, anthropogenic, environmental and climatic factors account for, respectively, 33, 27, 19 and 21 percent of the factors use in the models (Table 1).

Information on the ranking of importance of the 33 factors was extracted. From the 33 factors, 10 were the most used and ranked as being one of the five most essential factors for the models to map forest fire susceptibility. From the 10 factors, three were from the topographic category (slope, elevation and aspect), two were from the anthropogenic category (distance to roads and distance to residential areas), two were from the environmental category (land cover and NDVI) and three were from the climatic category (temperature, precipitation and wind speed). The three most important factors were distance to roads (67), elevation (67) and slope (68), which were ranked 46 and 41times, respectively, as being one of the five most essential factors in the models.

From the correlation table it can be determine that 24 of the factors that are used as inputs to map forest fire susceptibility are correlated. With five uncorrelated and four negatively correlated factors (Fig. 6).

The results are supported by the hierarchical clustering analysis. With three main clusters. The first cluster is composed by 24 factors (general factor category), the second cluster by four factors (specific factor category) and the third cluster by five factors (alternative factor category) (Fig S5). The first cluster is composed of factors such as distance to roads, elevation, slope etc. mainly general and correlated factors. This cluster also contains the 10 most used factors in forest fire susceptibility modelling which were identified previously. The second cluster is composed of fuel, population density, LST and life stock density, mainly more specific factors in comparison to cluster one and uncorrelated. The third cluster is composed of road density, WUI, distance to rails, shrubs, brought index mainly alternative factors of cluster one and negatively correlated.

Table 1 Factors ranking

Categories	Factors\ranking	1	2	3	4	5	Total	Used
Topographic	Slope	4	15	8	9	5	41	68
	Elevation	5	7	8	15	11	46	67
	Aspect	3	3	13	5	5	29	61
	Curvature			1		2	3	9
	Topographic Wet Index (TWI)		2	1		1	4	13
	Topographic Position Index (TPI)	1			1		2	4
Anthropogenic	Density roads			1			1	3
	Distance to roads	10	4	9	11	12	46	67
	Distance to rail				2	1	3	4
	Distance to residential areas	6	7	7	8	8	36	58
	Distance to agriculture			3	2		5	13
	Livestock density			1		1	2	3
	Population density	2	5		2		9	15
	Landuse	2	4		1	1	8	12
	Wild urban interface (WUI)	2			1		3	4
Environmental	Land cover	12	8	5	4	1	30	44
	Shrub		1	1			2	3
	Forest density	1			1		2	6
	normalize multiband drought index (NMDI)	1			1		2	4
	Normalize difference vegetation index (NDVI)	9	4	6	1	2	22	29
	Fuel	3	1				4	5
	Drought Index	1		1			2	4
	Soils			2	1	2	5	9
	Soil moisture	1		1			2	3
	Distance to Water		2	3	1	1	7	20
	Land surface temperature (LST)	2			1		3	5
Climatic	Temperature	6	6	3	3	6	24	35
	Maximum Temperature	1	1	2			4	7
	Mean Temperature	1	1	2		1	5	8
	Radiation	1		1	2		4	7
	Humidity		2	2	2	2	8	13
	Precipitation	10	4	8	6	4	32	50
	Wind speed			3		3	6	12	23

3.3.2 Models and accuracy

From the 94 articles, 63 reported on accuracy. The models that reported the highest and lowest accuracy are Bayesian Belief Network (98%) and Multiple Linear Regression (69%) (Table 2). Of the 33 models used in the past two decades 14 are above the average accuracy of 85%.

The most commonly used models were Random Forest (11), Linear Regression (8) and Analytic Hierarch Process (5). The five most commonly used models (RF, LR, AHP, ANN and MLR) account for 48% of models used in the past two decades. In period one
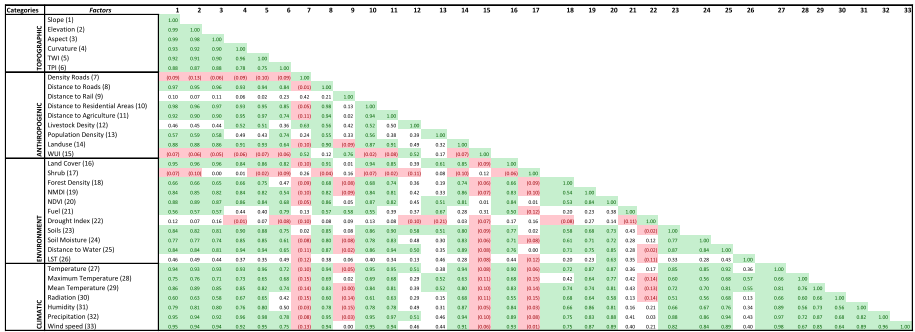
*Green = correlated, red = negatively correlated and white= uncorrelated

Fig. 6 Correlation Matrix

(2001–2011) the studies that reported accuracy were few in comparison with period two (2012–2021). In the past three years, there is an increase in the number and type of models used in mapping forest fire susceptibility (Fig. 7).

The Mann Whitney U test and Kruskal Wallis statistical tests indicated that accuracy do not differs based on study scale or county's level of development. The mean ranks for local and national scale studies were 30.47 and 39.23, respectively; the distributions in the two groups did not differed significantly (Mann–Whitney $U = 206.500$, $N_{local} = 52$ N $_{national} = 11$, $P = 0.150$ two-tailed). For countries level of development there was no statistically significant difference between the groups: H $(2) = 5.469$, $P = 0.065$, with a mean rank for low middle-income countries of 22.7, upper middle-income countries 35.27 and high-income countries 35.09.

The Independent t-test indicated that there is a statistically significant difference in accuracy based on the modelling method. The 32 machine learning algorithms ($M = 0.89000$, $SD = 0.070265$) compared to the 31 no machine learning algorithms ($M = 0.80326$, $SD = 0.075892$) demonstrated significantly better accuracy, $t (61) = 4.710$, $p = 0.000$.

4 Discussion

Based on the bibliometric analysis this study provides a more holistic perspective of forest fire susceptibility research in terms of the actors (countries and authors) involved. This has not been done before as previous reviews have focused on modelling methods, implementation and challenges (Naderpour et al. 2019; Jain et al. 2020; Abid 2021). The results presented highlights trends, but more importantly implementation and collaboration opportunities. A clear result is that there has been an increase in the use of models to map forest fire susceptibility since 2001. In many countries such research is now being conducted (Fig. 1). Figure 3, S2 and S3 illustrates that there are good international connections among countries and authors. It also illustrates the potential to enhance collaboration with and between countries in fire prone areas such as Brazil and South Africa. This gap is also illustrated by the results on co-authorship and implementation areas (Fig. 4). Co-authoring can greatly promote innovative studies and academic exchange (Tao et al. 2020). Fostering

Table 2 Average model accuracy

Number of models	Model name	Average accuracy
1	Bayesian belief network (BBNs)	0.980
1	Cascade generalization locally weighted learning (CG-LWL)	0.972
2	Bayes network (BN)	0.960
11	Random forest (RF)	0.930
1	Delphi method (MCDM-FL-WLC-ANP)	0.924
1	LogitBoost ensemble-based decision tree (LEDT)	0.920
1	Kernel logistic regression (KLR)	0.920
1	Particle swarm optimized neural fuzzy (PSO-NF)	0.916
2	Boosted regression tree (BRT)	0.896
1	VIKOR	0.895
1	Fuzzy membership value Index of Entropy (FMV-IOE)	0.890
1	Generalized additive mode (GAM)	0.878
1	Ensemble (EBF-LR)	0.864
1	Convolution neural network (CNN)	0.860
1	Ensemble (SVM-RF)	0.848
1	Binary logistic regression (BLR)	0.840
1	Fuzzy analytical hierarchical Process (FAHP)	0.835
1	Shanon entropy (SE)	0.830
1	Weight of evidence (WofE)	0.822
1	Multi-parametric weighted index (MPWI)	0.820
1	Evidentail belif function (EBF)	0.820
1	Fuzzy analytical network process (FANP)	0.819
4	Artificial neural network (ANN)	0.814
8	Logistic regression (LR)	0.813
1	Extreem gradient boosting (XGBoost)	0.804
1	Apriori and fuzzy C-means (Apriori-FCM)	0.800
5	Analytical hierarchical process (AHP)	0.788
1	Analytical network process (ANP)	0.783
1	Hybrid fire index (HFI)	0.770
2	Frequency ratio (FR)	0.765
2	Maximum entropy modelling (Maxent)	0.753
1	Generalized linear model (GLM)	0.750
3	Multiple linear regression (MLR)	0.689

research cooperation between currently unconnected researchers and places of research should be enhanced.

There is an increase in the use of models to map forest fire susceptibility around the globe (FigS. 4 and 7), due to an increase of forest fires (Tuyen et al. 2021) and improvements in technology (GIS and Remote Sensing). The results of this study indicate that there are fire prone areas where this type of research is not being implemented (Fig. 4). Especially, in Africa and Latin America known to be the most active fire areas (Costafreda-Aumedes et al. 2017). In Central America, for example, there was only one study
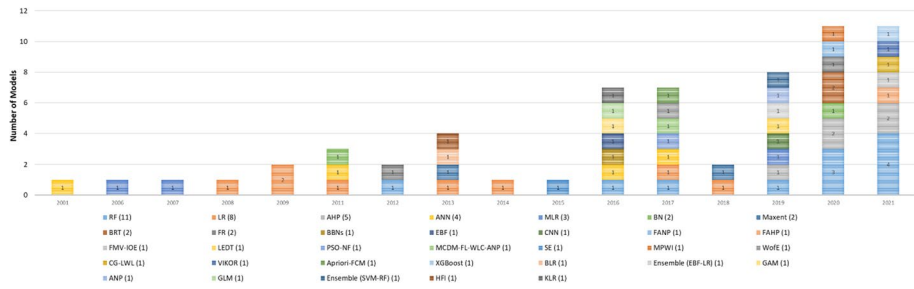
Fig. 7 Models used by year

conducted in Honduras in 2017 and this is a region that is highly affected by forest fires and where forest fire susceptibility data is lacking (Valdez et al. 2017). The lack of research in forest fire susceptibility mapping in Africa might be attributed to the notion that fires are an integral part of Africa's savanna ecosystem. However, forest fires pose significant constraint to sustainable economic development in Africa; thus, it is necessary to implement fire susceptibility research to improve fire management strategies (Dlamini 2011; Suryabhagavan et al. 2016). Moreover, with the recent increases in fire frequencies around the globe attributed to climate change it is urgent to conduct forest fire susceptibility research in fire prone regions, such as Latin America and Africa, to mitigate forest fires ecological, environmental and economic impacts (Faramarzi et al. 2021; Milanović et al. 2021; Sulova and Arsanjani 2021).

Over the past two decades, there have been three significant focus themes for forest fire susceptibility modelling research, factors, technology and models. A lot of the emphasis in recent years has been placed in modelling methods (Fig. 7). Our findings illustrate that factors and models used to map forest fire susceptibility varies across time and space. At the same scale, local, national or regional level, there is no common set of factors or models applied to map forest fire susceptibility. Environmental, climatic, topographic and anthropogenic differences explain this (Wang et al., 2020) but countries data availability and technical capacity also matters. Yet, our findings provide an overview of the most essential factors and models used in mapping forest fire susceptibility. From the 144 factors identified, 33 were used in more than three models. From these 33 factors, six out of the 10 most commonly used factors are from the topographic and climatic categories (Table 1). This might be attributed to data availability, as these factors are derived from freely available global databases in contrast to more specific anthropogenic factor, which need to be derived from local databases. Research in Iran and India have applied most of the 33 major factors identified (Fig S4). Iran and India are the countries where most forest fire susceptibility modelling studies were conducted. Within these two countries there are variations on the combination of factors used as inputs into the models to map forest fire susceptibility. Yet, the 33 factors identified provide a good overview of the base input factors used to model and map forest fire susceptibility, irrespective of the study region. Moreover, these identified factors are important predictors of forest fire susceptibility as depicted in Table 1.

The correlation and the hierarchical clustering analysis provide useful insights of the relationship among the 33 factors and which of these factors are used together when mapping forest fire susceptibility. The correlation analysis shows that from the 33 factors the 24 positively correlated factors (general factor category) have been used the most as inputs in the models to map forest fire susceptibility (Table 1). Negatively correlated factors (alternative factor category) have also been used but to a lesser

extent. For example, road density (alternative factor category) has been used three times in comparison to distance to roads (general factor category) which has been used 67 times in the models. It is essential to note that the four factors in the alternative category convey almost the same type of information as corresponding factors found in the general factor category. For instance, road density and distance to roads are factors that are usually derived from the same data layer or shrubs and landcover; however, based on the analysis general factors are favored over the alternative factors. The hierarchical clustering analysis furthered reduced the 24 positively correlated factors to 20 highly correlated factors that are used in different combinations when mapping forest fire susceptibility (Fig S5). These 20 correlated factors contain the ten most use factors in forest fire susceptibility models. These 10 factors have a correlation of greater or equal to 0.94 and are also ranked as one of the five most significant predictor of forest fire susceptibility. The high correlation of these factors means that these factors have been used together when mapping forest fire susceptibility. For example, if slope is used in a model as an input factor to map forest fire susceptibility, it is highly probable that elevation and aspect were also used as the correlation of these three factors is greater or equal to 0.98. The uncorrelated factors can be attributed to the low number of times these factors have been used five times or less. The low utilization of these specific factors is due to the difficulty of obtaining these factors from global databases since they contain more specific information when compared to the corresponding factors contained in the general factor category. For example, livestock density is more specific (specific factor category) in comparison to distance to agriculture which is more general (general factor category). Given the specific nature of these factors data availability is an issue. For instance, many of the specific anthropogenic and environmental factors identified were only used in one or two models. Some of these more specific factors, which were also significant predictors, are deforestation (Salame et al. 2012), agriculture machinery density (Martínez et al. 2009), land Tenure (Dlamini 2011), recreation areas (Vilar del Hoyo et al. 2011), unemployment (Oliveira et al. 2012), poverty (Qayum et al. 2020), biomass density (Prasad et al. 2008) and stand canopy closure (Güngöroğlu 2017). The utilization of more specific anthropogenic and environmental factors in forest fire susceptibility models is needed to provide more specific information to managers as to what are the vital factors that need to be address in order to mitigate forest fire susceptibility and improve natural resource management. The identification and decomposition of the 33 base factors, in terms correlation and clustering, provides critical information which when combined with the researcher's knowledge of the study area can improve factors selection to map forest fire susceptibility.

Model selection also needs to be considered when mapping forest fire susceptibility as model type affects accuracy (Malik et al. 2021; Tuyen et al. 2021). During the study period 2001–2021, 33 different models have been used globally to map forest fire susceptibility. However, 73% of the models have only been implemented in one study. Random Forest, Logistic Regression and Analytical Hierarchical Process are the models implemented the most. These models represent Machine Learning (RF), Statistical (LR) and Knowledge-based (AHP) methods, out of these RF had the highest reported average accuracy. This is supported by the Independent T test analysis which indicated that machine learning, ensemble and hybrid models have better accuracy than knowledge-based, statistical and other models. In recent years, hybrid and ensemble models are reporting better accuracies that machine learning models; thus, these models are being implemented in forest fire susceptibility mapping. However, the implementation

of these models is more complex (Naderpour et al. 2019); thus, for now, RF remains the most used method in modelling forest fire susceptibility. There was no significant difference in accuracy based on the scale of the study and the countries level of development because the majority of the studies used globally available databases to derive climatic, topographic, anthropogenic and environmental factors. This indicate that by using globally available datasets it is possible to conduct forest fire susceptibility modelling and obtain acceptable results in regions where local or national data limitations are an issue.

5 Conclusion

The utilization of models to map forest fire susceptibility is increasing. Therefore, this research provides vital insights on the actors (countries and autors) that are involved in this type of research, the main factors used in models in order to map forest fire susceptibility, and the models used. The identification of actors highlighted implementation and collaboration opportunities that can promote research cooperation between currently unconnected researchers and countries, which is urgently needed to mitigate the impacts of forest fires. The analysis of the base factors and models provide critical information which when combined with the researcher's knowledge of the study area can improve factor and model selection when mapping forest fire susceptibility. In addition, the results indicate that forest fire susceptibility mapping can be done in countries where there are local or national data limitations since the majority of the factors used, which are also essential predictors, can be derived from global databases. On the other hand, the general nature of these factors used limit the usefulness of the models at a local scale, which calls for the utilization of more specific factors that can aid managers to identify the specific drivers of forest fire susceptibility.

Declarations

Conflict of interest The authors declared that they have no conflict of interest.

References

Abedi Gheshlaghi H, Feizizadeh B, Blaschke T, Lakes T, Tajbar S (2021) Forest fire susceptibility modeling using hybrid approaches. Trans GIS 25(1):311–333

Abid F (2021) A survey of machine learning algorithms based forest fires prediction and detection systems. Fire Technol 57(2):559–590

Biresselioglu ME, Demir MH, Solak B, Kayacan A, Altinci S (2020) Investigating the trends in arctic research: the increasing role of social sciences and humanities. Sci Total Environ 729:139027

Bisquert M, Sánchez JM, Caselles V (2013) Modeling fire danger in Galicia and asturias (Spain) from MODIS images. Remote Sens 6(1):540–554

Briones-Bitar J, Carrión-Mero P, Montalván-Burbano N, Morante-Carballo F (2020) Rockfall research: a bibliometric analysis and future trends. Geosci 10(10):403. https://doi.org/10.3390/geosciences1010 0403

Carrión-Mero P, Montalván-Burbano N, Herrera-Narváez G, Morante-Carballo F (2021) Geodiversity and mining towards the development of Geotourism: a global perspective. Int J Design Nat Ecodynamics 16(2):191–201

Chuvieco E, Congalton RG (1989) Application of remote sensing and geographic information systems to forest fire hazard mapping. Remote Sens Environ 29(2):147–159

Costafreda-Aumedes S, Comas C, Vega-Garcia C (2017) Human-caused fire occurrence modelling in perspective: a review. Int J Wildl Fire 26(12):983. https://doi.org/10.1071/WF17026

Dang ATN, Kumar L, Reid M, Mutanga O. (2021) Fire danger assessment using geospatial modelling in Mekong delta, Vietnam: Effects on wetland resources. Remote Sensing Applications: Society and Environment. 21.

Dlamini WM (2011) Application of Bayesian networks for fire risk mapping using GIS and remote sensing data. GeoJournal 76(3):283–296

Eskandari S, Amiri M, Sãdhasivam N, Pourghasemi HR (2020) Comparison of new individual and hybrid machine learning algorithms for modeling and mapping fire hazard: a supplementary analysis of fire hazard in different counties of Golestan Province in Iran. Nat Hazards 104(1):305–327

Faramarzi H, Hosseini SM, Pourghasemi HR, Farnaghi M (2021) Forest fire spatial modelling using ordered weighted averaging multi-criteria evaluation. J Sci 67(2):87–100

Güngöroğlu C (2017) Determination of forest fire risk with fuzzy analytic hierarchy process and its mapping with the application of GIS: The case of Turkey/Çakırlar. Hum Ecol Risk Assess 23(2):388–406

Hawbaker TJ, Radeloff VC, Stewart SI, Hammer RB, Keuler NS, Clayton MK (2013) Human and biophysical influences on fire occurrence in the United States. Ecol Appl 23(3):565–582

Herrera-Franco G, Montalván-Burbano N, Carrión-Mero P, Bravo-Montero L (2021) Worldwide research on socio-hydrology: a bibliometric analysis. Water 13(9):1283. https://doi.org/10.3390/w13091283

Jain P, Coogan SCP, Subramanian SG, Crowley M, Taylor S, Flannigan MD (2020) A review of machine learning applications in wildfire science and management. Environ Rev 28(4):478–505

Kim SJ, Lim CH, Kim GS, Lee J, Geiger T, Rahmati O, Son Y, Lee WK (2019) Multi-temporal analysis of forest fire probability using socio-economic and environmental variables. Remote Sens 11(1):86. https://doi.org/10.3390/rs11010086

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 339:b2700

McDonald JH (2014) Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland. This web page contains the content of pages 157–164

Malik A, Rao MR, Puppala N, Koouri P, Thota VAK, Liu Q, Chiao S, Gao J (2021) Data-driven wildfire risk prediction in northern California. Atmosphere 12(1):19 (**English**)

Multitemporal analysis of NDVI and land surface temperature for modeling the probability of forest fire occurrence in central Mexico(2005) Proceedings of the Third International Workshop on the Analysis of Multi-Temporal Remote Sensing Images 2005.

Manzo-Delgado L, Sánchez-Colón S, Álvarez R (2009) Assessment of seasonal forest fire risk using NOAA-AVHRR: a case study in central Mexico. Int J Remote Sens 30(19):4991–5013

Martínez J, Vega-Garcia C, Chuvieco E (2009) Human-caused wildfire risk rating for prevention planning in Spain. J Environ Manage 90(2):1241–1252

Michael Y, Helman D, Glickman O, Gabay D, Brenner S, Lensky IM (2021) Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. Sci Total Environ 764:142844

Milanović S, Milanović SD, Marković N, Pamučar D, Gigović L, Kostić P (2021) Forest fire probability mapping in eastern serbia: Logistic regression versus random forest method. Forests 12(1):1–17

Montalván-Burbano N, Velastegui-Montoya A, Gurumendi-Noriega M, Morante-Carballo F, Adami M (2021) Worldwide research on land use and land cover in the amazon region. Sustainability 13(11):6039

Naderpour M, Rizeei HM, Khakzad N, Pradhan B (2019) Forest fire induced Natech risk assessment: a survey of geospatial technologies. Reliab Eng Syst Safety 191:106558

Oliveira S, Oehler F, San-Miguel-Ayanz J, Camia A, Pereira JMC (2012) Modeling spatial patterns of fire occurrence in mediterranean europe using multiple regression and random forest. For Ecol Manage 275:117–129

Poncet A, Courvoisier DS, Combescure C, Perneger TV (2016) Normality and sample size Do Not matter for the selection of an appropriate statistical test for two-group comparisons. Methodology 12(2):61–71

Pourghasemi HR, Gayen A, Lasaponara R, Tiefenbacher JP (2020) Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling. Environ Res 184:109321. https://doi.org/10.1016/j.envres.2020.109321

Pourghasemi HR, Mohammady M, Pradhan B (2012) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin. Iran Catena 97:71–84

Pourtaghi ZS, Pourghasemi HR, Aretano R, Semeraro T (2016) Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. Ecol Ind 64:72–84

Pourtaghi ZS, Pourghasemi HR, Rossi M (2015) Forest fire susceptibility mapping in the Minudasht forests, Golestan province, Iran. Environ Earth Sci 73(4):1515–1533

Prasad VK, Badarinath KVS, Eaturu A (2008) Biophysical and anthropogenic controls of forest fires in the Deccan Plateau, India. J Environ Manage 86(1):1–13

Qayum Abdul, Ahmad Firoz, Arya Rakesh, Singh Rajesh Kumar (2020) Predictive modeling of forest fire using geospatial tools and strategic allocation of resources: eForestFire. Stochastic Environ Res Risk Assessment 34(12):2259–2275 (**English**)

Razavi-Termeh SV, Sadeghi-Niaraki A, Choi SM (2020) Ubiquitous GIS-based forest fire susceptibility mapping using artificial intelligence methods. Remote Sensing 12(10):1689. https://doi.org/10.3390/rs12101689

Sakellariou S, Tampekis S, Samara F, Flannigan M, Jaeger D, Christopoulou O, Sfougaris A (2019) Determination of fire risk to assist fire management for insular areas: the case of a small Greek island. J Forest Res 30(2):589–601

Salame CW, Queiroz JCB, de Miranda RG, Amin MM (2012) Mapping the risk of burning in the brazilian amazon with the use of logistic regression and fuzzy inference. Math Geosci 44(3):241–256

Forest fire hazard modeling using hybrid AHP and fuzzy AHP methods using MODIS sensor (2012) International Geoscience and Remote Sensing Symposium (IGARSS).

Sulova A, Arsanjani JJ (2021) Exploratory analysis of driving force of wildfires in Australia: an application of machine learning within google earth engine. Remote Sensing 13(1):1–23

Suryabhagavan KV, Alemu M, Balakrishnan M (2016) Gis-based multi-criteria decision analysis for forest fire susceptibility mapping: a case study in Harenna forest, southwestern Ethiopia. Trop Ecol 57(1):33–43

Syphard AD, Radeloff VC, Keeley JE, Hawbaker TJ, Clayton MK, Stewart SI, Hammer RB (2007) Human influence on California fire regimes. Ecol Appl 17(5):1388–1402

Tao J, Qiu D, Yang F, Duan Z. (2020) A bibliometric analysis of human reliability research. Journal of Cleaner Production. 260.

Tehrany MS, Jones S, Shabani F, Martínez-Álvarez F, Tien BD (2019) A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. Theoret Appl Climatol 137(1–2):637–653

Tien Bui D, Bui QT, Nguyen QP, Pradhan B, Nampak H, Trinh PT (2017) A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. Agric for Meteorol 233:32–44

Tien Bui D, Hoang N-D, Samui P (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). J Environ Manage 237:476–487

Tuyen TT, Jaafari A, Yen HPH, Nguyen-Thoi T, Phong TV, Nguyen HD, Van Le H, Phuong TTM, Nguyen SH, Prakash I et al. (2021) Mapping forest fire susceptibility using spatially explicit ensemble models based on the locally weighted learning algorithm. Ecological Informatics. 63.

Valdez MC, Chang KT, Chen CF, Chiang SH, Santos JL (2017) Modelling the spatial variability of wildfire susceptibility in Honduras using remote sensing and geographical information systems. Geomat Nat Haz Risk 8(2):876–892

van Eck NJ, Waltman L (2017) Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics 111(2):1053–1070

Vasilakos C, Kalabokidis K, Hatzopoulos J, Matsinos I (2009) Identifying wildland fire ignition factors through sensitivity analysis of a neural network. Nat Hazards 50(1):125–143

Vilar del Hoyo L, Isabel MPM, Vega FJM (2011) Logistic regression models for human-caused wildfire risk estimation: Analysing the effect of the spatial accuracy in fire occurrence data. Eur J Forest Res 130(6):983–996

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

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;

2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;

3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;

4. use bots or other automated methods to access the content or redirect messages

5. override any security feature or exclusionary protocol; or

6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com