Applied Geography 151 (2023) 102867

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Modeling and prediction of fire occurrences along an elevational gradient in Western Himalayas

Somnath Bar [a], Bikash Ranjan Parida [a,*], Arvind Chandra Pandey [a], B. Uma Shankar [b], Pankaj Kumar [c], Santosh K. Panda [d], Mukunda Dev Behera [e]

[a] Department of Geoinformatics, School of Natural Resource Management, Central University of Jharkhand, Ranchi, 835222, India
[b] Machine Intelligence Unit, Indian Statistical Institute, Kolkata, 700108, India
[c] Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research Bhopal, India
[d] Department of Natural Resources and Environment and Institute of Agriculture, Natural Resources and Extension, University of Alaska Fairbanks, Fairbanks, AK, 99775, USA
[e] Centre for Oceans, Rivers, Atmosphere and Land Sciences CORAL, Indian Institute of Technology, Kharagpur, 721302, India

ARTICLE INFO

Handling Editor: E. Hunja Waithaka

Keywords:
Forest fire
Human activities
Climate
Machine learning
Random forest
Himalaya

ABSTRACT

Forest fires are the result of complex interactions among human, geographic and weather conditions. Climate change would alter the link between forest fire and the controlling factors. The objective of the study is to model the forest fire occurrences and quantify the contribution of explanatory geographic, climatic and anthropogenic variables using satellite-derived historical fire data (2003–2019) and machine learning classifiers over the western Himalaya, India. The climatic variables were derived from a regional Earth system model (ROM). Along with the key selected explanatory variables, the conditions of neighbouring (3 × 3) pixels were incorporated to account for the contribution from the surrounding area. Out of the selected classifiers, random forest recorded the most promising performance in k-fold cross-validation (f2-score = 0.95 and f1-score = 0.94) as well as in the final model validation (f2-score = 0.85 and f1-score = 0.84). The elevation and mean neighbour elevation exhibited the highest influence (8.18% and 6.72%, respectively) in forest fire occurrences followed by near-surface temperatures (4.65–5.78%). We predicted the forest fire susceptibility [0, 1] for 2030, 2040 and 2050 using the future climate projections. The predicted map can be useful to plan effective fire management strategies to minimize damage to the forest ecosystem.

1. Introduction

The forest system plays a fundamental role in the global biogeochemical cycle, predominately in the carbon cycle (Shen et al., 2019). Forest fire is an integral and inevitable component of terrestrial ecosystems and has a significant contribution to ecosystem functionality, land atmospheric interaction, and energy flux (Bowman et al., 2009). In the last few decades, the influence of forest fires has been amplified in shaping the environment and atmosphere (Shi et al., 2021). Especially, forest ecosystems are increasingly threatened by natural and human fire activities. Wildland or forest fire has significantly contributed to the budgets of atmospheric aerosol and trace gases (Akagi et al., 2011), which alter the local and regional air quality. Moreover, this changing atmospheric composition has influenced to change the short-term weather as well as climate (Ramanathan & Carmichael, 2008). Studies indicated that climate change may lead to more intense and frequent fire activities (Flannigan et al., 2013). Warm, windy and dry conditions favour fires over the Northwest, USA, and such conditions are becoming more frequent under climate change (Goss et al., 2020; Halofsky, Peterson, & Harvey, 2020). The recent climatic change has played an important role in increasing forest fire intensity over the Amazon forest, mainly in the drought years. Consequently, the lower precipitation and tropical North Atlantic Ocean induced warming have caused severe and extensive forest fires across the Amazon (Chen et al., 2017; Jolly et al., 2015; Marengo et al., 2008). The majority of the global fire activities are found along the tropics, and savannas are the most affected land cover. Tropical forest fires in south and Central America, Africa, and Southeast Asia (Maurin et al., 2014; Pearson, Brown, Murray, & Sidman, 2017) and Asian crop residue burning (Sahu & Sheel, 2014) account for around 90% of the world's fires. In every year, the global average burnt area size

* Corresponding author.
E-mail address: bikash.parida@cuj.ac.in (B.R. Parida).

https://doi.org/10.1016/j.apgeog.2022.102867
Received 23 June 2022; Received in revised form 13 December 2022; Accepted 27 December 2022
Available online 30 December 2022

is ~423 Mha which equates to the area of the European Union (Giglio, Boschetti, Roy, Humber, & Justice, 2018). This large fire activity not only degrades the forest covers but causes loss of lives and properties. Therefore, understating the underlining factors and predicting future forest fire occurrences is an increasingly important task.

Forest fires are becoming a dominant natural disaster and threat to forest ecosystems across the globe (Bowman et al., 2009). The ignition and spread of a forest fire is a result of a complex interaction among climate, topography (elevation, slope, aspect), forest type and human activities (Tonini et al., 2020). The elevation significantly controlled other local physio-climatic conditions from a high ridge to a gentle valley and the steeper slope aided to burn up-slope very fast (Patton & Coen, 2004). Additionally, the climatic parameters (i.e. temperature, precipitation, humidity, solar radiation, wind etc.), and biophysical condition of vegetation (Fraction of Absorbed Photosynthetically Active Radiation, Fraction of green Vegetation Cover, Leaf Area Index, Chlorophyll content, etc.) have significant control over the fire regime (size, pattern, frequency and intensity) of an ecosystem (Estes, Knapp, Skinner, Miller, & Preisler, 2017; Liu, Ballantyne, & Cooper, 2019; Marlon et al., 2008). Like the weather component, temperature positively accelerates the forest fire intensity (Bar, Parida, Roberts, et al., 2021) but precipitation immediately suppresses the fire activities. However, in the long run, it encourages high fire activities, through vegetation and fuel build-up (Earl & Simmonds, 2018; Marlon et al., 2008). It is a fact that the regional climatic condition is substantially regulated by large-scale climatic oscillations, i.e. El-Niño–Southern Oscillation (ENSO), and teleconnection (Mason, Hamlington, Hamlington, Matt Jolly, & Hoffman, 2017). Apart from the climatic and environmental factors, human activities are responsible for ~90% of fire ignition, but the prevalent environment creates the indispensable condition which accelerates the intensity of fire (Bar, Parida, Roberts, et al., 2021; Taylor, 2010). Consequently, the quantification of contributing geo-environmental and human factors is essential to understand the underlining causes and predicting future fire activities. Machine learning (ML) algorithms are becoming popular in forest fire detection, occurrence modeling and prediction studies (Abid, 2020; Bar, Parida, & Pandey, 2020a; Collins, Griffioen, Newell, & Mellor, 2018).

Effective management and prevention techniques are essential to control forest fire activities, which could be possible through the appropriate prediction of spatial fire probability (Abid, 2020). Fire susceptibility estimates the probability of fire occurrence over space and time. Regional forest fire susceptibility is regulated by many factors and has typically nonlinear relations with complex characteristics. Hence, it is challenging to develop a promising forest fire predictable model (Ngoc Thach et al., 2018). Several approaches were developed in forest fire susceptibility modeling from physical process-based methods to statistical, and ML-based modeling (Abid, 2020; Chang et al., 2013; Han, Ryu, Chi, & Yeon, 2003; Tonini et al., 2020). The data-driven ML algorithms have also drawn wide attention for decades (Abid, 2020). The machine learning algorithms, such as random forest (RF) (Tonini et al., 2020; G. Zhang, Wang, & Liu, 2019), support vector machine (SVM) (Sakr et al., 2010), artificial neural networks (ANN), multilayer perceptron neural network (Dimuccio et al., 2011; Zhang, Wang, & Liu, 2021), kernel logistic regression (Hong, Pradhan, Xu, & Tien Bui, 2015), naïve Bayes (Avitabile et al., 2016) and gradient boosting decision tree (Dimuccio et al., 2011) were utilized in forest fire susceptibility modeling. Among the algorithms, the RF model demonstrated a highly promising accuracy in forest fire susceptibility modeling (Collins et al., 2018; Ramo, García, Rodríguez, & Chuvieco, 2018; Tonini et al., 2020).

India is one of the forest fire-prone regions in Southeast Asia, specifically the deciduous forests of central and south India, and forests in the west and east Himalayas (Bar, Parida, Pandey, & Kumar, 2022; Bar, Parida, Pandey, & Panda, 2022; Pratap Srivastava, 2013; Vadrevu et al., 2013). In recent decades, the northwestern Himalaya experienced devastating forest fire activities, mostly during the hot and dry pre-monsoon (March to June) period (Babu. et al., 2016; Dobriyal & Bijalwan, 2017). The burn area increased at a rate of 72.94 km^2 year^{-1} over Uttarakhand and Himachal Pradesh between 2001 and 2019 (Bar, Parida, Roberts, et al., 2021). Some of the limited studies demonstrated the forest fire risk modeling based on geospatial techniques (Babu. et al., 2016; Kumar, Meenakshi, Das Bairagi, Vandana, & Kumar, 2015), Analytic Hierarchy Process (AHP) and fuzzy AHP (Tiwari, Shoab, & Dixit, 2021) in the parts of western Himalaya. The remote sensing and geospatial information system demonstrated a relevant role in forest fire monitoring and assessment (Bar et al., 2020b; Chuvieco et al., 2018; Giglio, Schroeder, & Justice, 2016; Reddy et al., 2019). The Moderate Resolution Imaging Spectroradiometer (MODIS) derived global fire products (e.g., MOD14A1, MOD14A2, MYD14CMQ, MOD14CMQ, MCD64CMQ, MCD14ML, and MCD64A1) are one of the reliable fire dataset. From our knowledge, the present study is the first of its kind that modeled forest fire occurrences and predicted future fire susceptibility. The overarching objectives of the study are to (i) model the forest fire occurrences over western Himalayan, using remote sensing and regional climate model-derived spatial data and ML algorithms, (ii) quantify the contribution of geographic, climatic and anthropogenic variables on forest fires, and (iii) predict the fire susceptibility for 2030, 2040 and 2050.

2. Study area

The study area covers parts of the western Himalaya, Himachal Pradesh and Uttarakhand states of India, which is a fragile mountain ecosystem and rich in biodiversity. Importantly, fire is a frequent human and naturally driven phenomenon in the Himalayas (Dobriyal & Bijalwan, 2017). The major forest types in this mountain ecosystem are evergreen broadleaf and needle leaf forest, and deciduous broadleaf forests (Reddy, Jha, Diwakar, & Dadhwal, 2015). The elevation ranges from about 173 m to 7764 m (above mean sea level), and the topography is characterized by steep mountain ridges and narrow valleys (Fig. 1). The climatic condition ranges from tropical to temperate (Mal & Singh, 2014; Parida, Pandey, & Patel, 2020), and the warmer summer or pre-monsoon (March to June) is the indispensable period for fire activities (Bar, Parida, Roberts, et al., 2021). The annual precipitation varies from 600 to 2000 mm which is mostly found in the monsoon season (July to September). Maximum temperature ranges from 15 °C to 40 °C, the warmest temperature found between March and June (pre--monsoon) (V. Kumar, Shanu, & Jahangeer, 2017; Mohd Wani, Sarda, & Jain, 2017). The pre-monsoonal forest fire burnt area has significantly increased over the last two decades (2001–2019) over this region (Bar, Parida, Roberts, et al., 2021). Therefore, probability or susceptibility based fire occurrence modeling and future prediction are essential for this fragile mountainous ecosystem.

3. Data and methods

The forest fire occurrence was modeled using the co-occurring geographic, climatic and anthropogenic variables, and the methedological workflow presented in Fig. 2. The whole dataset has been divided into two parts, static geographical information (e.g., topographic, locational, forest type including anthropogenic) and dynamic climatic variables. The detailed description of the data was explained in the following sections and, and listed in Table 1.

3.1. Fire data

The MODIS data is a reliable source of fire products over the last two decades. The present study has adopted a MODIS-derived collection 6 vector fire dataset (MCD14ML v0061) from 2003 to 2019. The MCD14ML data was produced using both AQUA and TERRA thermal information, processed by the National Aeronautics and Space Administration (NASA) Science Computing Facility (SCF) of the University Of Maryland (UMD). MOD14 and MYD14 swath products flagged the active

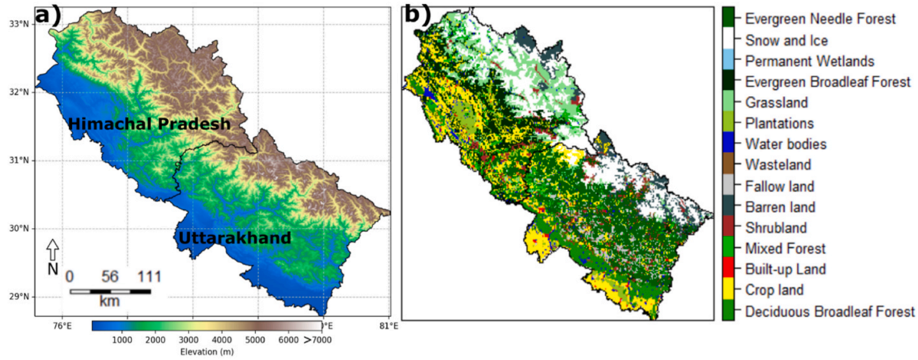

Fig. 1. (a) Elevation map of the study area (Uttarakhand and Himachal Pradesh) derived from Advanced Land Observing Satellite (ALOS) World 3D (30 m) digital surface model (DSM) and (b) Land Use land cover map, extracted from Roy et al. (2015).

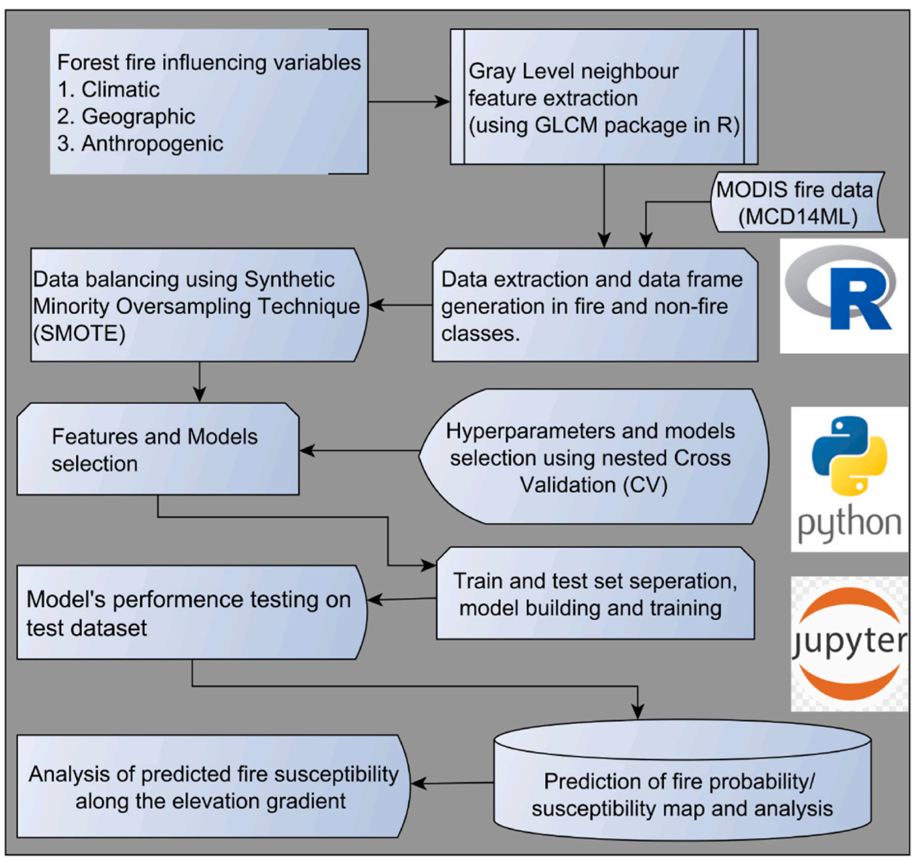

Fig. 2. The methodological workflow of forest fire modeling and prediction over western Himalaya.

fire (MCD14ML), at centre of 1 km pixel, through thermal anomaly (Giglio, Descloitres, Justice, & Kaufman, 2003; Justice et al., 2010). Compared to MCD14ML collection 5 and 5.1 (earlier versions), this fire product reduces false alarms and uncertainties (Giglio, Randerson, & van der Werf, 2013). The vector spatial fire point locations are distributed from Fire Information for Resource Management System (FIRMS). The present study has considered fire points having a confidence level of more than 80%. The MODIS burn area data product, i.e., MCD64A1 was also used to illustrate the annual burn areas of the study area from 2003 to 2019.

3.2. Climatic variables

The climatic variables, namely near surface maximum, minimum and average temperature (°C), accumulated precipitation (mm/day), relative (%) and specific humidity (kg/kg) at 1000 and 850 hPa, solar radiation (W/m^2) and wind vector at 10 m height at ~20 km spatial resolution are used to model the fire occurrences of western Himalaya (Table 1). These climatic data was retrieved on a daily scale from a regional Earth system model (ROM) (Mishra et al., 2021; Sein et al., 2015). ROM has been forced with a global model MPI_ESM_LR using historical, moderate (Representative Concentration Pathway 4.5 (RCP4.5)) and high emission scenario (RCP8.5) forcing for period 1950–2100. The model was simulated over the CORDEX-South-Asia domain (Mishra et al., 2021) at horizontal grid resolution of 0.22° × 0.22°. The ROM derived climatic variables (maximum temperature, minimum temperature and precipitation) have exhibited a promising association (82–92%) with the India Meteorological Department (IMD) climatic data.

Table 1
A list of data products used to accomplish the forest fire modelling and prediction over western Himalaya.

Data	Product ID	Characteristics	Source
Active fire	MCD14ML v0061	Daily and point	Fire Information for Resource Management System (FIRMS)
Burn area	MCD64A1 v006	Monthly and 500m	Land Processes Distributed Active Archive Center (LP DAAC)
Climatic Variables (near surface (2 m) maximum and minimum temperature (°C), accumulated precipitation (mm/day), relative (%) and specific humidity (kg/kg) at 1000 and 850 hPa, solar radiation (W/m2) and wind vector at 10m)		Daily and ~20 km	Regionally coupled atmosphere-ocean-sea ice-marine biogeochemistry model (ROM)
Digital Surface Model (DSM)	AW3D30	~30 m	Advanced Land Observing Satellite (ALOS)
Population density	GPWv411	~1 km × 1 km, demi-decadal	SEDAC (2018)
Global Human Modification	gHM	~1 km × 1 km, available in 2016 only	Kennedy, Oakleaf, Theobald, Baruch-Mordo, and Kiesecker (2019b)

3.3. Topographic and land use land cover (LULC)

The topographic variables significantly influence fire ignition and its spreading over the landscape. The study area mostly covers the undulating terrain of the western Himalayas. The topographic condition and land cover information are used as static geographical information. We used a 30 m digital surface model (DSM) retrieved from Advanced Land Observing Satellite (ALOS; AW3D30) data (Takaku, Futamura, Iijima, Tadono, & Shimada, 2007). Surface elevation (m), slope (°), aspect (°) and roughness were computed from the ALOS-based DSM. The forest area and type (including shrub and grassland) information were retrieved from a decadal land cover land use product of 2005 (Roy et al., 2015), with a spatial resolution of 100 m.

3.4. Anthropogenic variables

Anthropogenic activities have a major contribution, mainly in igniting forest fires (Bar, Parida, Roberts, et al., 2021; Puri, Areendran, Raj, Mazumdar, & Joshi, 2011). Therefore, it is indispensable to consider the human footprint data in forest fire modeling. A gridded population density data (GPWv411) generated through the extrapolation of census data, available every 5 years (2005 − 2020) in ~1 km was used (SEDAC, 2018). The global Human Modification (gHM) intensity index is a cumulative measure of human settlement, agriculture, transportation (major, minor, and two-track roads; railroads), mining and energy production and electrical infrastructure (power lines, nighttime lights) intensity (Kennedy, Oakleaf, Theobald, Baruch-Mordo, & Kiesecker, 2019a). The gHM of 2016 (~1 km) was employed to include the various human modification related effects on forest fire over the western Himalaya.

3.5. Co-occurring neighbour feature extraction

According to the first law of Geography, the nearer things are more closely associated than remote things (Tobler, 1970). While ignition of a pixel and fire spread is not only dependent upon the weather geographic and vegetation conditions in that pixel/region, but also depends on the conditions of neighbouring pixels/regions (Zhang et al., 2019). The pixel-based machine learning classifiers assumed fire or non-fire is a function of the different explanatory factors of that pixel only. However, it does not take into account the state of the explanatory variables of neighbouring/associated regions/pixels in a particular fire or non-fire pixel Looking into this gap, the present study has extracted the neighbour information i.e., mean and entropy. A 3 × 3 kernel has been run over the images to compute the mean and entropy of the central pixel. Along the margin of the image or boundary regions, the co-occurring information were retrieved using the available neighbour pixels. The gray level co-occurrence matrixes (GLCM) package of R (Lu & Batistella, 2005) was used to derive the gray level neighbour information for selected variables.

3.6. Machine learning algorithms

The study has adopted numbers ML classifiers, namely Logistic Regression (LR), Gaussian Naive Bayes (GNB), Hist Gradient Boosting (HGB), Extreme Gradient Boosting (XGB) and RF. These ML classifiers are commonly used to model processes and events in the geoscience discipline, specifically in forest fire modeling studies (Abid, 2020; Acharya et al., 2021; Chang et al., 2013; Feng, Zhao, Chen, & Zhang, 2020; Tonini et al., 2020; Zhang et al., 2019), especially forest fire detection and prediction. The model parameters are briefly explained in the next section.

3.7. Features selection

Initially, the key forest fire influencing variables were selected from the literature (Abid, 2020; Bar, Parida, & Uma Shankar, 2021; Dimuccio et al., 2011; Sakr, Elhajj, Mitri, & Wejinya, 2010; Tiwari et al., 2021). The climatic (12 variables), geographic (5 variables), anthropogenic (2 variables), and their 2 Gy level neighbour variables (i.e., mean and entropy) were used as influencing agents in the ignition and spread of forest fire. Along with these variables, two geo-locational variables (i.e., latitude and longitude) were also employed as covariates to account for the spatial distribution of forest fire occurrences. The main 19 variables and their 38 (19 × 2 = 38) neighbour occurrence information and 2 locational variables (total 59 variables) were extracted for every pixel. The variable names and their code names are present in Table S1. The feature selection was carried out using two methods i.e., recursive feature elimination (RFE) and Mean Decreasing Accuracy (MDA) of the RF model. The RFE is an efficient algorithm in feature selection that attempts to reduce the collinearity of the training dataset (Thilagam & V.S, 2007). The RF model is used as the base model in RFE. The RFE was performed 18 times (starting with 6 features and 3 more were included with each iteration) and from a total of 59 features, 50 features were selected as explanatory variables for forest fire occurrence modeling.

3.8. Model and parameters selection

A double or nested Cross-Validation (CV) hypertuning approach was adopted to select the hyperparameters of the aforementioned models (Cawley & Talbot, 2010). The simple hyperparameter optimization could overfit the dataset, while the nested CV reduces the bias and aid in selecting the hyperparameters and model (Cawley & Talbot, 2010). The nested CV method uses an iterative pair of nested loops, in the outer loop, the hyperparameters are adjusted to optimise a model selection criterion (i.e., model selection). In the inner loop, the parameters are set to optimise a training criterion (model fitting). Here, we used 10-fold CV in the outer loop and 5-fold CV in the inner loop. The best combination of the hyperparameters of the models and the average and standard deviation (SD) of the models' performance matrices were presented in Table 2.

Table 2
Nested k-fold cross-validation model-wise hyper-parameters and corresponding model performance metrics.

Model	Optimum-parameters	Accuracy matrixes (mean and standard deviation)
Logistic Regression	max_iter = 200, penalty = '12', solver = 'lbfgs'	f2-score = 0.818 (±0.003), f1-score = 0.777 (±0.003), recall = 0.848 (±0.003), precision = 0.717 (±0.004) and Accuracy = 0.767 (±0.002)
Gaussian Naive Bayes	var_smoothing = 0.23101297	f2-score = 0.837 (±0.004), f1-score = 0.748 (±0.004), recall = 0.909 (±0.005), precision = 0.651(±0.004) and Accuracy = 0.711 (±0.004)
Hist Gradient Boosting Classifier	l2_regularization = 1, learning_rate = 0.5, max_iter = 500, max_depth = 30	f2-score = 0.908 (±0.004), f1-score = 0.904 (±0.005), recall = 0.911 (±0.005), precision = 0.898 (±0.007) and Accuracy = 0.906 (±0.005)
XGBoost Classifier	n_estimators = 2000, max_depth = 10, learning_rate = 0.2	f2-score = 0.943 (±0.003), f1-score = 0.942 (±0.003), recall = 0.943 (±0.004), precision = 0.943 (±0.004) and Accuracy = 0.940 (±0.004)
Random Forest Classifier	n_estimators = 2000, criterion = 'entropy', max_features = 8, max_depth = 40.	f2-score = 0.958 (±0.002), f1-score = 0.940 (±0.002), recall = 0.970 (±0.003), precision = 0.912 (±0.003) and Accuracy = 0.940 (±0.002)

3.9. Model building

After getting the optimum hyperparamerters, and the average accuracy matrix from the nested CV, the final model was built. The dataset was split into two parts, the training dataset was from 2003 to 2018, and the trained model was tested over 2019's dataset. As the weather conditions and human activities were different in 2020–2021, due to the COVID-19 related lockdown (Parida, Bar, Kaskaoutis, et al., 2021; Parida, Bar, Roberts, et al., 2021; Parida, Bar, Singh, et al., 2021), we didn't consider data from 2020 to train or test the model. The extracted training dataset of fire occurrence (y) classes (i.e., fire; 1 and non-fire; 0) were in a considerable high-class imbalance, which could create bias in favour of the majority class and reduce the overall model performance. To eradicate this class imbalance and get an optimally balanced training dataset, this study has employed Synthetic Minority Over-sampling Technique (SMOTE) (Blagus & Lusa, 2013). SMOTE did not generate any new information for the dataset. In the present study, the majority and minority classes of the training dataset were non-fire (n = 4136126) and fire (n = 10414) respectively, which resulted in a balanced training dataset (i.e., n = 52450 in each class). The contribution of the explanatory variables (i.e., feature importance) retrieved from the mean decreasing accuracy (MDA) of RF (Hong Han, Guo, & Yu, 2016).

The accuracy of nested CV and final model validation was assessed using precision, recall, f2–score, f1–score, Overall Accuracy, and area under the curve (AUC) (Goutte & Gaussier, 2005). These statistical measures were present in Supplement – 2.

4. Results

4.1. Historical distribution of forest fire

Fig. 3 presents the spatial distribution of forest fire events and inter-annual burn area variability during the study period. The spatial distribution of fire counts in 5 × 5 km grid cells over 18 years suggests the highest fire frequency (i.e., 35–40) along the southern and central part of the study area, specifically over Uttarakhand. The spatial distribution of fire events and frequency revealed a locational association (Fig. 3a) with a significant interannual variability (SD = 568.58 km^2). The average pre-monsoonal burn area of this region was 748.25 km^2. The highest burn area was observed in 2012 (1781.71 km^2) and 2016 (1753.57 km^2), followed by 2009 and 2019 (Fig. 3b).

4.2. Performance of the models in nested cross-validation

The performances of the models were assessed through a number of accuracy measures in the outer loop (i.e., 10–folds CV). The highest and consistent performance was found in RF classifier (i.e., f2–score = 0.958 (±0.002), f1–score = 0.940 (±0.002), recall = 0.970 (±0.003), precision = 0.912 (±0.004) and OA = 0.940 (±0.002)). The best combination of the hyperparameters is given in Table 2. GNB and LR classifiers exhibited the lowest performance based on the accuracy measures. The XGB and HGB classifiers performance ranked after RF (i.e., the f2–score were 0.943 and 0.908, respectively). Based on the performance of models (Table 2), the study selected RF classifier for the future prediction of forest fires.

4.3. Relative importance of the features

Fig. 4 shows the relative feature importance (%) of the selected variables. Where, the elevation (8.62%), mean neighbour elevation (i.e., Elev_m; 6.71%) and temperatures at 2-m height (i.e., Temp2m, Tmin2m and Tmax2m; 5.80–4.78%) demonstrated the highest contribution in

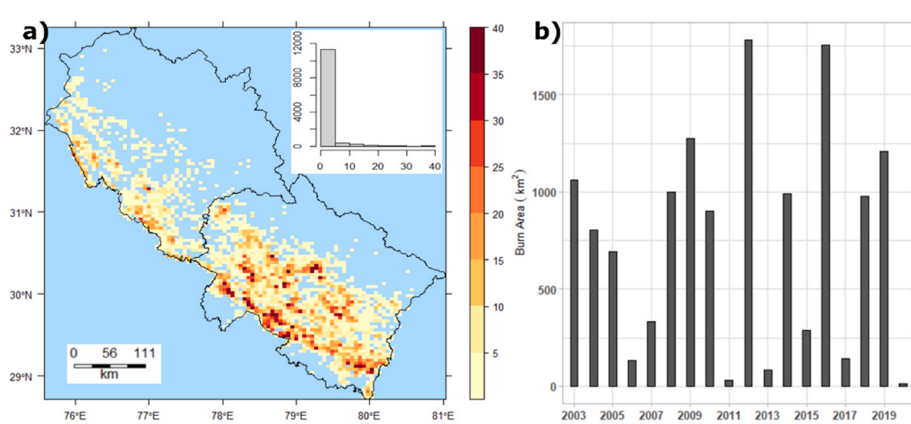

Fig. 3. (a) Fire counts in 5 × 5 km grid over 2003–2020 according to MCD14ML active fire data and (b) variability of pre-monsoonal total burn area (MCD64A1) over Uttarakhand and Himachal Pradesh.

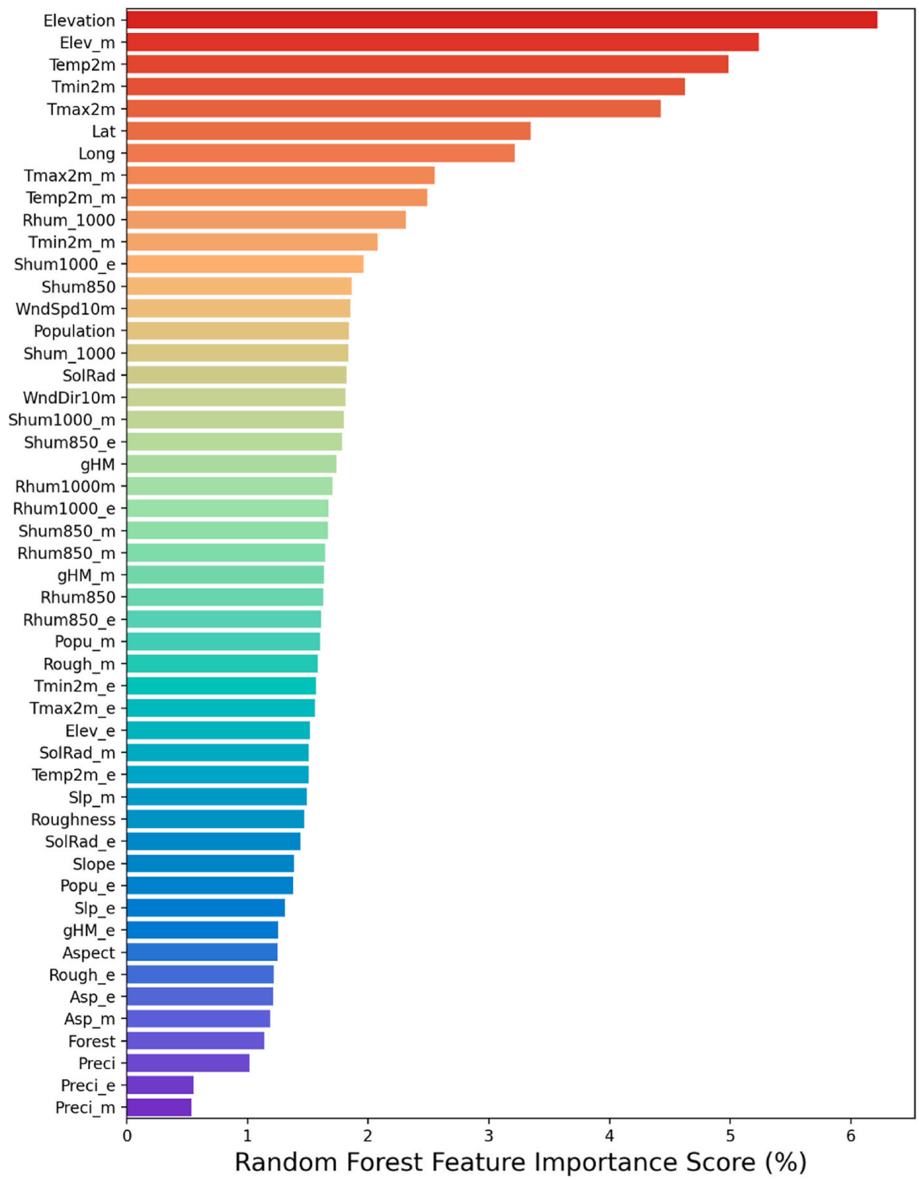

Fig. 4. Random Forest derived feature importance (%) in forest fire modeling.

forest fire occurrences over western Himalaya. The mountainous topography of this region significantly controls other local physio-climatic conditions, from a high ridge to a gentle valley. Here, elevation shows the highest contribution to forest fire occurrence than the near-surface temperatures. Followed by the locational covariates i.e., latitude and longitude (3.96% and 3.82%, respectively), Tmax2m_m, Temp2m_m and Rhum1000 (i.e., 2.74%, 2.50% and 2.34%, respectively) and population (2.06%) and others variables (Fig. 4). The lowest contributing variables were the precipitation and aspect variables. However, orders of feature importance slightly vary among models (RF, XGB, HGB and LR). Still, as RF demonstrated the highest and most consistent performance in nested CV, the study only considered the feature importance of RF.

4.4. Validation of the final model

The reported accuracy measures of the validations were presented in Table 3. With respect to both of the classes (i.e., fire and non-fire), RF demonstrated the highest accuracy (f2–score = 0.85, f1–score = 0.84, recall = 0.85, precision = 0.84, OA = 0.85 and AUC = 0.84), followed by XGB (OA = 0.83 and AUC = 0.82) and HGB (OA = 0.78 and AUC = 0.77). The GNB classifier recorded the lowest agreement in respect of all the accuracy measures (Table 3). The spatial agreements between the

Table 3
Models' performance metrics on the tested dataset of 2019 fire occurrences.

Model	Accuracy metrics					
	F2-score	F1-score	Recall	Precision	Overall Accuracy (OA)	AUC
Logistic Regression	0.83	0.77	0.88	0.69	0.70	0.83
Gaussian Naive Bayes	0.72	0.70	0.73	0.67	0.67	0.70
Hist-Gradient Boosting Classifier	0.76	0.77	0.76	0.78	0.78	0.77
XGBoost Classifier	0.81	0.82	0.86	0.83	0.83	0.82
Random Forest Classifier	0.85	0.84	0.85	0.84	0.85	0.84

actual fire points and predicted forest fire susceptibility based on LR, GNB, HGB, XGB and RF were presented in Fig. 5a–e, respectively. These susceptibility maps suggest the major forest fire clusters are located along the southern slope of the study area, especially over Uttarakhand. The actual active fire points of 2019 were mostly located in/and around higher forest fire probability (close to 1) regions.

4.5. Prediction of forest fire susceptibility

This study also predicted the forest fire susceptibility for the years 2030, 2040 and 2050 using the climatic scenarios of RCP4.5 and RCP8.5. The geographic and anthropogenic variables were kept as static variables. The RF derived projected forest fire susceptibility maps were presented in Fig. 6. Both scenarios revealed a higher probability (close to 1) of forest fire along the southern slope of the study area in the selected time period, which could be attributed to increasing near-surface temperature. According to the RCP8.5 scenario, the higher intensity of fire activities would gradually extend to the middle and northern parts of the study area in the next decades. Both of the scenarios suggested a slightly decreasing trend of fire activities (in 2050) over Himachal Pradesh and an increasing trend in Uttarakhand (Fig. 6). We also analysed the predicted fire susceptibility in 2030, 2040 and 2050 along with the change of elevation, our model projected the highest probability (>0.8) of fire at elevation <2500 m, and a lower probability of fire at higher elevation regions (Fig. 7).

5. Discussion

An appropriate prediction of forest fire susceptibility is a prerequisite to control, manage and reduce fire events and damages. The present study demonstrated the potential utilities of ML algorithms to forest fire modeling and prediction using large numbers of climatic, geographic and anthropogenic explanatory variables. Long-term historical active fire observation data (2003–2018) was used to train the models of fire occurrences. The key findings of the study indicate that the RF classifier performed comparatively better in nested CV (f2–score = 0.96 and f1–score = 0.94), as well as in model validation over 2019 (f2–score = 0.85 and f1–score = 0.84), out of the five selected classifier. As the study area is characterized by mountainous undulating topography, the elevation, Elev_m and near-surface temperatures (Temp2m, Tmin2m and Tmax2m) were revealed as the highest contributing variables in forest fire occurrence. Whereas previous studies over the western Himalayan region concluded that most of the fire activities are confined within ~2000 m of elevation where near surface-temperature is comparatively higher (Babu, Roy, & Prasad, 2016; Bar, Parida, & Uma Shankar, 2021; Roy, 2005). Typically, the forest fire of this region is location-specific, especially over the southern slope, therefore, the present study found a substantial influence of geo-locational variables in fire occurrence. The neighbour conditions (gray level information: mean and entropy) of the fire and non-fire pixels showed a significant control in fire occurrence, especially the neighbour temperature i.e., Tmax2m_m and Temp2m_m. The anthropogenic variables (i.e., population and gHM) confirmed a moderate contribution in forest fire incidence over the study area. The precipitation and their neighbour (i.e., Preci_m and Preci_e) variables accounted for a lower contribution in forest fire incidences of western Himalaya, because the present study considered only the pre-monsoon season (March to June) which is mostly the dry season, very fewer amounts of precipitation occur over the main fire-affected parts of the region. Both bottom-up (topography, forest type) and top-down (climatic and human activities) controls of the forest fire depicted a significant contribution in forest fire but specifically, the bottom-up i.e., elevation revealed the highest controls over forest fire of this region. A similar finding was reported by Gill & Taylor (2009) in California, USA which indicated fire regimes were strongly influenced by bottom-up factors (i.e., elevation gradient). The forest fire susceptibility over 2030, 2040 and 2050 was predicted using RCP4.5 (moderate) and RCP8.5 (high emission) climate projections. This indicates higher fire activities along the southern slopes of the study area and the fire

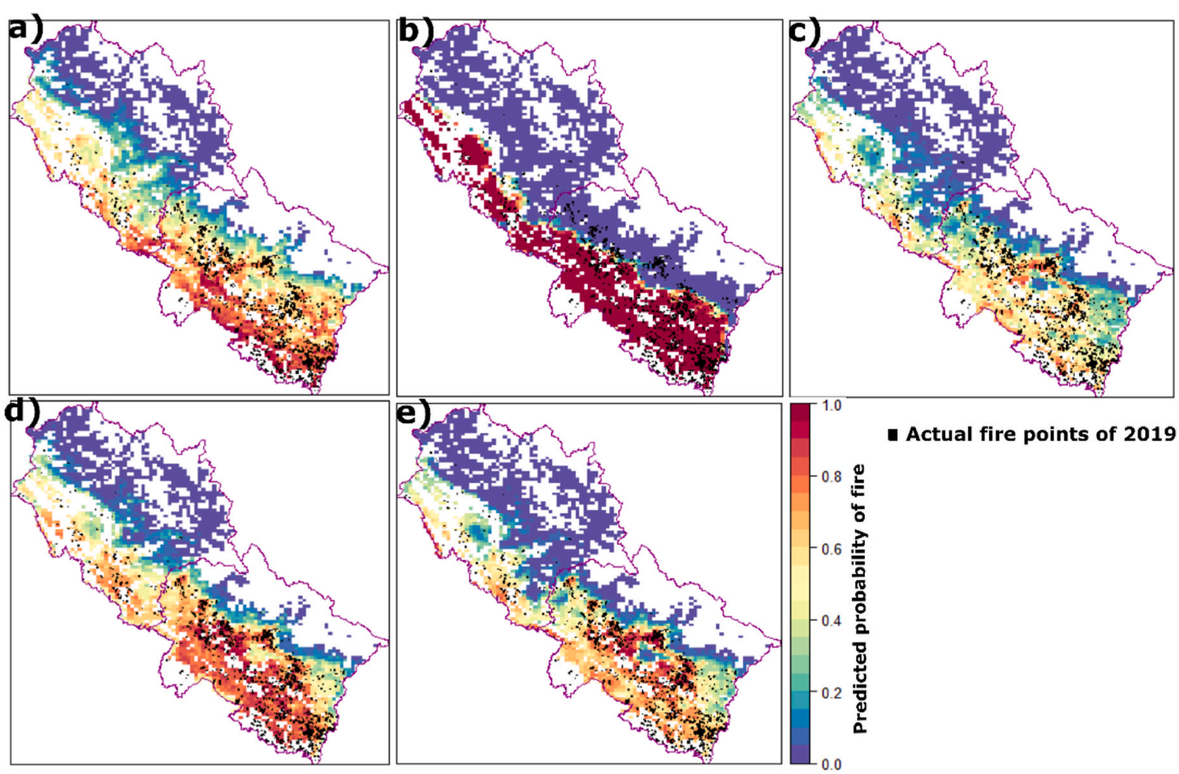

Fig. 5. (a) Logistic regression, (b) Gaussian Naive Bayes, (c) Hist Gradient Boosting, (d) XGBoost and (e) Random Forest derived forest fire susceptibility (0–1) map, where 0 and 1 indicate no-fire and high probability to fire, respectively. The actual fire active points of 2019 were overlaid.

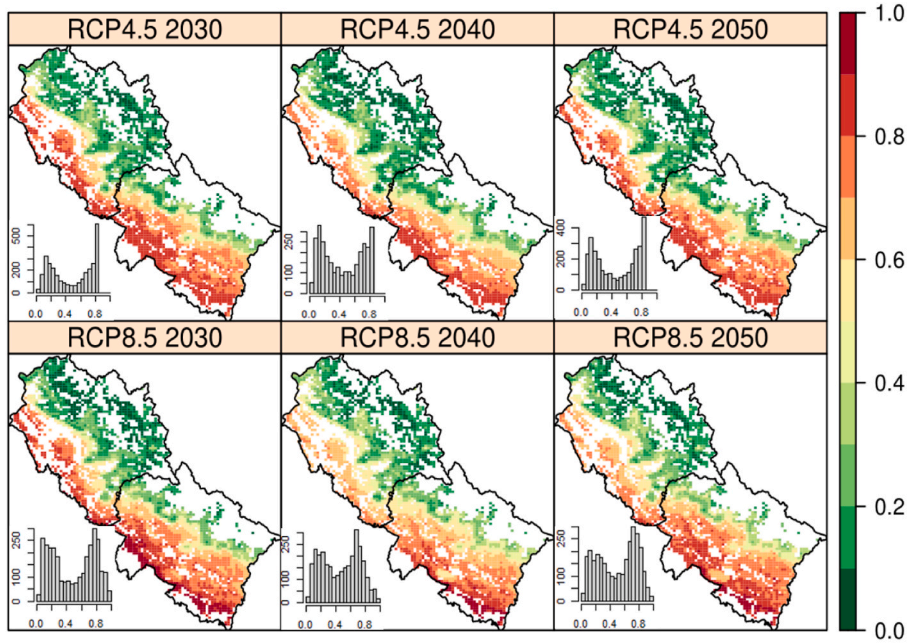

Fig. 6. Random Forest derived predicted susceptibility of forest fire and corresponding histogram over 2030, 2040 and 2050 using RCP4.5 and RCP8.5 climatic scenarios.

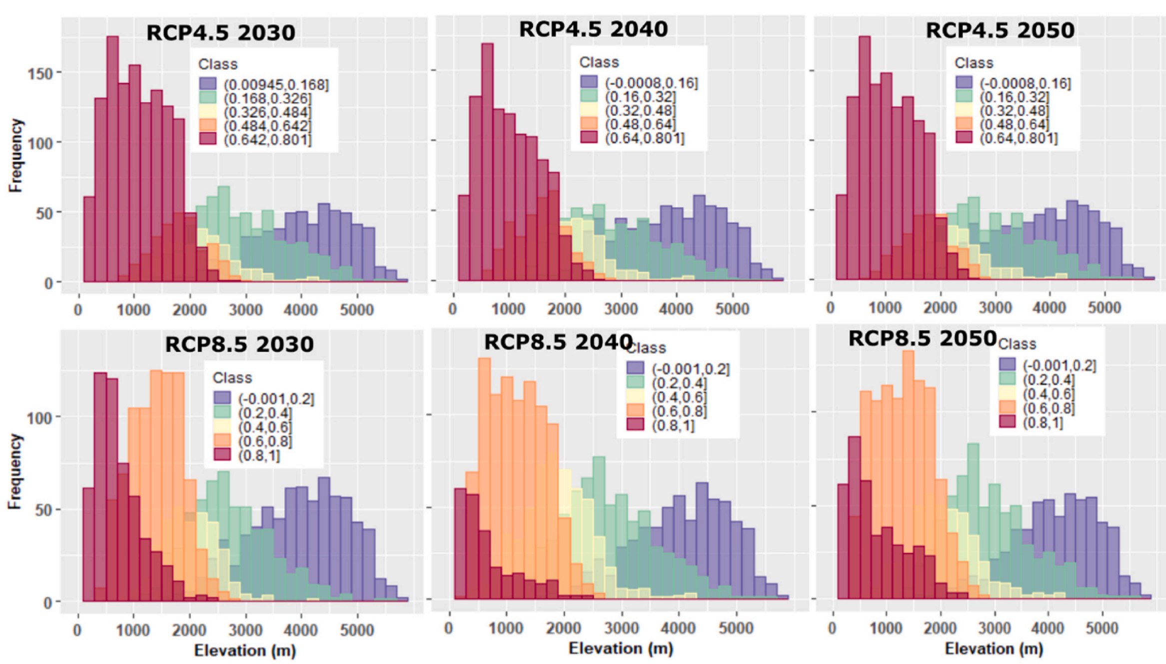

Fig. 7. Predicted forest fire susceptibility along the gradient of elevation (m) over Uttarakhand and Himachal Pradesh. The class value and corresponding colour code present the fire susceptibility alongside elevation.

intensity would gradually extend to middle and northern higher elevated parts, specifically over Uttarakhand. The elevation gradient showed that higher fire susceptibility pixels would be mostly over <2500m elevation regions due to sufficient fuel loads, comparatively higher temperature and dryness and human activities. Moreover, Babu et al. (2016) and Bar, Parida, and Uma Shankar (2021) identified similar factors that were responsible for higher fire activities in western Himalaya.

In this study, we adopted a nested CV approach to hyperparameters optimization and model selection which reduces the overfitting probability of CV as well as in the parameters selection method (Cawley & Talbot, 2010). We incorporated the neighbourhood information (i.e., mean and entropy) along with the main explanatory variables, to account for the influence of the neighbour pixel's condition in forest fire occurrence and spread. As per the performance matrix, the RF was the comparatively more promising model than XGB, HGB, GNB and LR. The previous machine learning-based forest fire probability studies also reported a superior performance of RF classifiers (Collins et al., 2018; Pourtaghi, Pourghasemi, Aretano, & Semeraro, 2016; Tonini et al., 2020; Zhang et al., 2019) over different parts of the globe. Among the selected models (i.e., RF, XGB, HGB, GNB and LR), GNB had the lowest performance (i.e. f2–score = 0.72, f1–score = 0.70 and AUC = 0.70), which might be due to the collinearity, feature selection, data volume and data type (Zhang, 2004). We analysed the collinearity among the climatic, geographic and anthropogenic variables, but if we drop the variables based on collinearity then a good number of important variables (near surface-temperatures, relative and absolute humidity, elevation, population density, human modification, among others) will be out of the model, which directly reduces the performance of the models.

The remote sensing derived geographic, human, and climatic; i.e., regional earth system model-derived variables were employed to model the fire occurrences over the western Himalaya. We used the active fire product of MODIS, version 6 (MCD14ML). The MCD14ML product is generated from both Terra and Aqua thermal channels (in ~1 km), the product would miss the fire spots not actively burning during the satellite overpass. In the process of forest fire prediction, this study assumed the geographic including geo-locational variables and anthropogenic variables as static variables where the future predicted geographic and anthropogenic variables need to incorporate while forest fire prediction to get more robust predicted fire susceptibility. Another significant shortcoming of the study is that the future predicted forest fire susceptibility is heavily reliant on the projected climatic scenarios, which are based on a few assumptions and have significant uncertainty. Apart from all this, the present study could be used to navigate the short-term prediction of forest fires and their control and prevention by the forest authority.

6. Conclusions

The historical forest fire occurrence from MODIS was modeled as a function of selected geographic, human and climatic (ROM derived) variables for the western Himalaya. The study considered both geo-environmental conditions and human activities to understand forest fire occurrences (i.e., fire and non-fire) using ML models or classifiers. The hyper-parameters and the selection of the best classifier were carried out using a nested CV grid searching method, which reduces the overfitting probability. RF appeared as the most promising classifier in nested CV as well as in the final model validation followed by XGB, HGB, LR and GNB. The elevation and the Elev_m recorded the highest contribution in forest fire occurrences. Followed by the near-surface temperatures, geo-locational covariates and mean neighbour temperatures and relative humidity. A promising association (up to 85%) between actual and predicted fire was found. As well, the spatial associations of predicted susceptibility and actual forest fire of 2019 exhibited a close association. The study demonstrated the potential use of satellite, reanalysis and regional climate model-derived data in forest fire modeling and prediction over a mountainous region. The temporal availability of geographical and anthropogenic data is limited which can reduce the prediction accuracy of the forest fire. The predicted future fire susceptibility maps could aid to navigate the probable fire occurrence spots to control and manage forest fires. Nevertheless, it has been suggested that forest management authorities may start implementing fuel breaks along the southern slope as well as around high-value resources.

CRediT authorship contribution statement

S. Bar, B.R. Parida, A.C. Pandey: Conceptualization, Investigation, Methodology, Software, Data Analysis, Visualization, writing– original draft, review and editing. B.U. Shankar, P. Kumar: Data support, Data Analysis, writing–review and editing. S.K. Panda and M.D. Behera: writing–review and editing, All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

Authors declare no potential conflict of interest.

Acknowledgement

Authors are thankful to the Earth Data of NASA for MCD14ML, MCD64A1 fire datasets. First author thanks to UGC for assisting NET-JRF fellowship (3289/(SC NET-JAN 2017) for doctoral research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apgeog.2022.102867.

References

Abid, F. (2020). A survey of machine learning algorithms based forest fires prediction and detection systems. *Fire Technology, 57*(2), 559–590. https://doi.org/10.1007/s10694-020-01056-z

Acharya, P., Barik, G., Gayen, B. K., Bar, S., Maiti, A., Sarkar, A., et al. (2021). Revisiting the levels of Aerosol Optical Depth in south-southeast Asia, Europe and USA amid the COVID-19 pandemic using satellite observations. *Environmental Research, 193* (October 2020), Article 110514. https://doi.org/10.1016/j.envres.2020.110514

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., et al. (2011). Emission factors for open and domestic biomass burning for use in atmospheric models. *Atmospheric Chemistry and Physics, 11*(9), 4039–4072. https://doi.org/10.5194/acp-11-4039-2011

Avitabile, V., Herold, M., Heuvelink, G. B. M., Lewis, S. L., Phillips, O. L., Asner, G. P., et al. (2016). An integrated pan-tropical biomass map using multiple reference datasets. *Global Change Biology, 22*(4), 1406–1420. https://doi.org/10.1111/gcb.13139

Babu, K. V. S., Roy, A., & Prasad, P. R. (2016). Forest fire risk modeling in Uttarakhand Himalaya using TERRA satellite datasets. *European Journal of Remote Sensing, 49*(1), 381–395. https://doi.org/10.5721/EuJRS20164921

Bar, S., Parida, B. R., & Pandey, A. C. (2020a). *Landsat-8 and Sentinel-2 based Forest fire burn area mapping using machine learning algorithms on GEE cloud platform over Uttarakhand*. Western Himalaya: Remote Sensing Applications: Society and Environment, Article 100324. https://doi.org/10.1016/j.rsase.2020.100324, 18 (March.

Bar, S., Parida, B. R., Pandey, A. C., & Kumar, N. (2022). Pixel-based long-term (2001–2020) estimations of forest fire emissions over the Himalaya. *Remote Sensing, 14*(21), 5302. https://doi.org/10.3390/rs14215302

Bar, S., Parida, B. R., Pandey, A. C., & Panda, S. K. (2022). Changing forest fire regime in relation to climatic conditions over western and eastern Himalaya, India. *In Handbook of Himalayan Ecosystems and Sustainability, 1*, 281–300. https://doi.org/10.1201/9781003268383-19. CRC Press.

Bar, S., Parida, B. R., Roberts, G., Pandey, A. C., Acharya, P., & Dash, J. (2021). Spatio-temporal characterization of landscape fire in relation to anthropogenic activity and climatic variability over the Western Himalaya, India. *GIScience and Remote Sensing, 58*(2), 281–299. https://doi.org/10.1080/15481603.2021.1879495

Bar, S., Parida, B. R., & Uma Shankar, B. (2021). Unfolding the contribution of environmental and anthropogenic variables in forest fire over western Himalayan fire regime. *IEEE International India Geoscience and Remote Sensing Symposium (InGARSS)*, 557–560. https://doi.org/10.1109/InGARSS51564.2021.9792002, 2021.

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics, 14*(1), 1–16. https://doi.org/10.1186/1471-2105-14-106/FIGURES/7

Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., et al. (2009). Fire in the earth system. *Science, 324*(5926), 481–484. https://doi.org/10.1126/science.1163886

Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research, 11*, 2079–2107.

Chang, Y., Zhu, Z., Bu, R., Chen, H., Feng, Y., Li, Y., et al. (2013). Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China. *Landscape Ecology, 28*(10), 1989–2004. https://doi.org/10.1007/s10980-013-9935-4

Chen, Y., Morton, D. C., Andela, N., van der Werf, G. R., Giglio, L., & Randerson, J. T. (2017). A pan-tropical cascade of fire driven by El Niño/Southern Oscillation. *Nature Climate Change, 7*(12), 906–911. https://doi.org/10.1038/s41558-017-0014-8

Chuvieco, E., Lizundia-Loiola, J., Pettinari, M. L., Ramo, R., Padilla, M., Tansey, K., et al. (2018). Generation and analysis of a new global burned area product based on MODIS 250& thinsp;m reflectance bands and thermal anomalies. *Earth System Science Data, 10*(4), 2015–2031. https://doi.org/10.5194/essd-10-2015-2018

Collins, L., Griffioen, P., Newell, G., & Mellor, A. (2018). The utility of Random Forests for wildfire severity mapping. *Remote Sensing of Environment, 216*(June), 374–384. https://doi.org/10.1016/j.rse.2018.07.005

Dimuccio, L. A., Ferreira, R., Cunha, L., & Campar De Almeida, A. (2011). Regional forest-fire susceptibility analysis in central Portugal using a probabilistic ratings procedure and artificial neural network weights assignment. *International Journal of Wildland Fire, 20*(6), 776–791. https://doi.org/10.1071/WF09083

Dimuccio, L. A., Ferreira, R., Cunha, L., & Campar de Almeida, A. (2011). Regional forest-fire susceptibility analysis in central Portugal using a probabilistic ratings procedure and artificial neural network weights assignment. *International Journal of Wildland Fire, 20*(6), 776. https://doi.org/10.1071/WF09083

Dobriyal, M. J. R., & Bijalwan, A. (2017). Forest fire in western Himalayas of India: A review. *New York Science Journal, 10*(6). https://doi.org/10.7537/marsnys100617.06

Earl, N., & Simmonds, I. (2018). Spatial and temporal variability and trends in 2001–2016 global fire activity. *Journal of Geophysical Research: Atmospheres, 123*(5), 2524–2536. https://doi.org/10.1002/2017JD027749

Estes, B. L., Knapp, E. E., Skinner, C. N., Miller, J. D., & Preisler, H. K. (2017). Factors influencing fire severity under moderate burning conditions in the Klamath Mountains, northern California, USA. *Ecosphere, 8*(5), Article e01794. https://doi.org/10.1002/ECS2.1794

Feng, Z., Zhao, Z., Chen, S., & Zhang, H. (2020). *Research on multi-factor forest fire prediction model using machine learning method in China*.

Flannigan, M., Cantin, A. S., de Groot, W. J., Wotton, M., Newbery, A., & Gowman, L. M. (2013). Global wildland fire season severity in the 21st century. *Forest Ecology and Management, 294*, 54–61. https://doi.org/10.1016/j.foreco.2012.10.022

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., & Justice, C. O. (2018). The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sensing of Environment, 217*, 72–85. https://doi.org/10.1016/j.rse.2018.08.005

Giglio, L., Descloitres, J., Justice, C. O., & Kaufman, Y. J. (2003). An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment, 87*(2–3), 273–282. https://doi.org/10.1016/S0034-4257(03)00184-6

Giglio, L., Randerson, J. T., & van der Werf, G. R. (2013). Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *Journal of Geophysical Research: Biogeosciences, 118*(1), 317–328. https://doi.org/10.1002/jgrg.20042

Giglio, L., Schroeder, W., & Justice, C. O. (2016). *Remote Sensing of Environment the collection 6 MODIS active fi re detection algorithm and fi re products* (Vol. 178, pp. 31–41).

Gill, L., & Taylor, A. H. (2009). Top-down and bottom-up controls on fire Regimes along an elevational gradient on the East Slope of the Sierra Nevada, California, USA. *Fire Ecology, 5*(3), 57–75. https://doi.org/10.4996/fireecology.0503057.

Goss, M., Swain, D. L., Abatzoglou, J. T., Sarhadi, A., Kolden, C. A., Williams, A. P., et al. (2020). Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. *Environmental Research Letters, 15*(9), Article 094016. https://doi.org/10.1088/1748-9326/ab83a7

Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European conference on information retrieval* (pp. 345–359). Springer. https://doi.org/10.1007/978-3-540-31865-1_25.

Halofsky, J. E., Peterson, D. L., & Harvey, B. J. (2020). Changing wildfire, changing forests: The effects of climate change on fire regimes and vegetation in the pacific northwest, USA. *Fire Ecology, 16*(1), 4.

Han, H., Guo, X., & Yu, H. (2016). *Variable selection using mean decrease accuracy and mean decrease gini based on random forest* (pp. 219–224). 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS). https://doi.org/10.1109/ICSESS.2016.7883053, 0.

Han, J. G., Ryu, K. H., Chi, K. H., & Yeon, Y. K. (2003). Statistics based predictive geo-spatial data mining: Forest fire Hazardous area mapping application. *Lecture Notes in Computer Science, 2642*, 370–381. https://doi.org/10.1007/3-540-36901-5_38. Issue January 2014.

Hong, H., Pradhan, B., Xu, C., & Tien Bui, D. (2015). Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena, 133*, 266–281. https://doi.org/10.1016/j.catena.2015.05.019

Jolly, W. M., Cochrane, M. A., Freeborn, P. H., Holden, Z. A., Brown, T. J., Williamson, G. J., et al. (2015). Climate-induced variations in global wildfire danger from 1979 to 2013. *Nature Communications, 6*(1), 7537. https://doi.org/10.1038/ncomms8537

Justice, C. O., Giglio, L., Roy, D., Boschetti, L., Csiszar, I., Davies, D., et al. (2010). MODIS-derived global fire products. In *Remote sensing and digital image processing* (Vol. 11, pp. 661–679). Springer International Publishing. https://doi.org/10.1007/978-1-4419-6749-7_29

Kennedy, C. M., Oakleaf, J. R., Theobald, D. M., Baruch-Mordo, S., & Kiesecker, J. (2019a). Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Global Change Biology, 25*(3), 811–826. https://doi.org/10.1111/gcb.14549

Kennedy, C. M., Oakleaf, J. R., Theobald, D. M., Baruch-Mordo, S., & Kiesecker, J. (2019b). Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Global Change Biology, 25*(3), 811–826. https://doi.org/10.1111/GCB.14549

Kumar, S., Meenakshi, Das Bairagi, G., Vandana, & Kumar, A. (2015). Identifying triggers for forest fire and assessing fire susceptibility of forests in Indian western Himalaya using geospatial techniques. *Natural Hazards, 78*(1), 203–217. https://doi.org/10.1007/s11069-015-1710-1

Kumar, V., Shanu, & Jahangeer. (2017). Statistical distribution of rainfall in Uttarakhand, India. *Applied Water Science, 7*(8), 4765–4776. https://doi.org/10.1007/s13201-017-0586-5

Liu, Z., Ballantyne, A. P., & Cooper, L. A. (2019). Biophysical feedback of global forest fires on surface temperature. *Nature Communications, 10*(1), 1–9. https://doi.org/10.1038/s41467-018-08237-z

Lu, D., & Batistella, M. (2005). Exploring TM image texture and its relationships with biomass estimation in Rondônia, Brazilian Amazon. *Acta Amazonica, 35*(2), 249–257. https://doi.org/10.1590/S0044-59672005000200015

Mal, S., & Singh, R. B. (2014). *Changing glacial lakes and associated outburst floods risks in Nanda Devi Biosphere Reserve, Indian Himalaya*. IAHS-AISH Proceedings and Reports.

Marengo, J. A., Nobre, C. A., Tomasella, J., Oyama, M. D., Sampaio de Oliveira, G., de Oliveira, R., et al. (2008). The drought of amazonia in 2005. *Journal of Climate, 21*(3), 495–516. https://doi.org/10.1175/2007JCLI1600.1

Marlon, J. R., Bartlein, P. J., Carcaillet, C., Gavin, D. G., Harrison, S. P., Higuera, P. E., et al. (2008). Climate and human influences on global biomass burning over the past two millennia. *Nature Geoscience, 1*(10), 697–702. https://doi.org/10.1038/ngeo313

Mason, S. A., Hamlington, P. E., Hamlington, B. D., Matt Jolly, W., & Hoffman, C. M. (2017). Effects of climate oscillations on wildland fire potential in the continental United States. *Geophysical Research Letters, 44*(13), 7002–7010. https://doi.org/10.1002/2017GL074111

Maurin, O., Davies, T. J., Burrows, J. E., Daru, B. H., Yessoufou, K., Muasya, A. M., … Bond, W. J. (2014). Savanna fire and the origins of the 'underground forests' of Africa. *New Phytologist, 204*(1), 201–214. https://doi.org/10.1111/nph.12936.

Mishra, A. K., Kumar, P., Dubey, A. K., Javed, A., Saharwardi, M. S., Sein, D. V., et al. (2021). Impact of horizontal resolution on monsoon precipitation for CORDEX-south Asia: A regional earth system model assessment. *Atmospheric Research, 259*, Article 105681. https://doi.org/10.1016/j.atmosres.2021.105681

Mohd Wani, J., Sarda, V. K., & Jain, S. K. (2017). Assessment of trends and variability of rainfall and temperature for the district of Mandi in Himachal Pradesh, India. *Slovak Journal of Civil Engineering.* https://doi.org/10.1515/sjce-2017-0014

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

Parida, B. R., Bar, S., Kaskaoutis, D., Pandey, A. C., Polade, S. D., & Goswami, S. (2021a). Impact of COVID-19 induced lockdown on land surface temperature, aerosol, and urban heat in Europe and North America. *Sustainable Cities and Society, 75*, Article 103336. https://doi.org/10.1016/j.scs.2021.103336

Parida, B. R., Bar, S., Roberts, G., Mandal, S. P., Pandey, A. C., Kumar, M., et al. (2021b). Improvement in air quality and its impact on land surface temperature in major urban areas across India during the first lockdown of the pandemic. *Environmental Research, 199*(May), Article 111280. https://doi.org/10.1016/j.envres.2021.111280

Parida, B. R., Bar, S., Singh, N., Oinam, B., Pandey, A. C., & Kumar, M. (2021c). A short-term decline in anthropogenic emission of CO 2 in India due to COVID-19 confinement. *Progress in Physical Geography: Earth and Environment, 45*(4), 471–487. https://doi.org/10.1177/0309133320966741

Parida, B. R., Pandey, A. C., & Patel, N. R. (2020). Greening and browning trends of vegetation in India and their responses to climatic and non-climatic drivers. *Climate, 8*(8), 92. https://doi.org/10.3390/cli8080092

Patton, E. G., & Coen, J. L. (2004). *WRF-fire : A coupled atmosphere-fire module for WRF*. Boulder, CO: Preprints of Joint MM5/Weather Research and Forecasting Model Users' Workshop. *22–25 June, NCAR, 221–223*, 3 http://www.mmm.ucar.edu/mm5/workshop/ws04/Session9/Patton Edward.pdf.

Pearson, T. R. H. H., Brown, S., Murray, L., & Sidman, G. (2017). Greenhouse gas emissions from tropical forest degradation: an underestimated source. *Carbon Balance and Management, 12*(1), 3. https://doi.org/10.1186/s13021-017-0072-2.

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

Pratap Srivastava, A. G. (2013). Forest fires in India : Regional and temporal analyses. *Journal of Tropical Forest Science, 25*(2), 151–156, 25(2), 228–239 https://www.jstor.org/stable/23617038.

Puri, K., Areendran, G., Raj, K., Mazumdar, S., & Joshi, P. K. (2011). Forest fire risk assessment in parts of Northeast India using geospatial tools. *Journal of Forestry Research.* https://doi.org/10.1007/s11676-011-0206-4

Ramanathan, V., & Carmichael, G. (2008). Global and regional climate changes due to black carbon. *Nature Geoscience, 1*(4), 221–227. https://doi.org/10.1038/ngeo156

Ramo, R., García, M., Rodríguez, D., & Chuvieco, E. (2018). A data mining approach for global burned area mapping. *International Journal of Applied Earth Observation and Geoinformation, 73*(April), 39–51. https://doi.org/10.1016/j.jag.2018.05.027

Reddy, C. S., Bird, N. G., Sreelakshmi, S., Manikandan, T. M., Asra, M., Krishna, P. H., et al. (2019). Identification and characterization of spatio-temporal hotspots of forest fires in South Asia. *Environmental Monitoring and Assessment.* https://doi.org/10.1007/s10661-019-7695-6

Reddy, C. S., Jha, C. S., Diwakar, P. G., & Dadhwal, V. K. (2015). Nationwide classification of forest types of India using remote sensing and GIS. *Environmental Monitoring and Assessment, 187*(12), 777. https://doi.org/10.1007/s10661-015-4990-8

Roy, P. S. (2005). Forest fire and degradation assessment using satellite remote sensing and geographic information system. *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology*, 361–400.

Roy, P., Roy, A., Joshi, P., Kale, M., Srivastava, V., Srivastava, S., et al. (2015). Development of decadal (1985–1995–2005) land use and land cover database for India. *Remote Sensing, 7*(3), 2401–2430. https://doi.org/10.3390/rs70302401

Sahu, L. K., & Sheel, V. (2014). Spatio-temporal variation of biomass burning sources over South and Southeast Asia. *Journal of Atmospheric Chemistry, 71*(1), 1–19. https://doi.org/10.1007/s10874-013-9275-4.

Sakr, G. E., Elhajj, I. H., Mitri, G., & Wejinya, U. C. (2010). *Artificial intelligence for forest fire prediction* (pp. 1311–1316). IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM. https://doi.org/10.1109/AIM.2010.5695809. September 2015.

SEDAC. (2018). Population density, v4.11: Gridded population of the world (GPW), v4 | SEDAC. https://doi.org/10.7927/H49C6VHW.

Sein, D. V., Mikolajewicz, U., Gröger, M., Fast, I., Cabos, W., Pinto, J. G., et al. (2015). Regionally coupled atmosphere-ocean-sea ice-marine biogeochemistry model ROM: 1. Description and validation. *Journal of Advances in Modeling Earth Systems, 7*(1), 268–304. https://doi.org/10.1002/2014MS000357

Shen, H., Tao, S., Chen, Y., Odman, M. T., Zou, Y., Huang, Y., et al. (2019). Global fire forecasts using both large-scale climate indices and local meteorological parameters. *Global Biogeochemical Cycles, 33*(8), 1129–1145. https://doi.org/10.1029/2019GB006180

Shi, Y., Gong, S., Zang, S., Zhao, Y., Wang, W., Lv, Z., et al. (2021). High-resolution and multi-year estimation of emissions from open biomass burning in Northeast China during 2001–2017. *Journal of Cleaner Production, 310*(March), Article 127496. https://doi.org/10.1016/j.jclepro.2021.127496

Takaku, J., Futamura, N., Iijima, T., Tadono, T., & Shimada, M. (2007). *High resolution DSM generation from ALOS PRISM*. International Geoscience and Remote Sensing Symposium (IGARSS). https://doi.org/10.1109/IGARSS.2007.4423215

Taylor, D. (2010). Biomass burning, humans and climate change in Southeast Asia. *Biodiversity & Conservation, 19*(4), 1025–1042. https://doi.org/10.1007/s10531-009-9756-6

Thilagam, P. S., & A, V. S. (2007). *Semantic partition based association rule mining across multiple databases using abstraction* (pp. 81–86). Sixth International Conference on Machine Learning and Applications. https://doi.org/10.1109/ICMLA.2007.44. *ICMLA 2007)*.

Tiwari, A., Shoab, M., & Dixit, A. (2021). GIS-Based forest fire susceptibility modeling in Pauri Garhwal, India: A comparative assessment of frequency ratio, analytic hierarchy process and fuzzy modeling techniques. *Natural Hazards, 105*(2), 1189–1230. https://doi.org/10.1007/s11069-020-04351-8

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(sup1), 234–240. https://doi.org/10.2307/143141

Tonini, M., D'andrea, M., Biondi, G., Esposti, S. D., Trucchia, A., & Fiorucci, P. (2020). A machine learning-based approach for wildfire susceptibility mapping. The case study of the liguria region in Italy. *Geosciences, 10*(3), 1–18. https://doi.org/10.3390/geosciences10030105

Vadrevu, K. P., Csiszar, I., Ellicott, E., Giglio, L., Badarinath, K. V. S., Vermote, E., et al. (2013). Hotspot analysis of vegetation fires and intensity in the Indian region. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6*(1), 224–238. https://doi.org/10.1109/JSTARS.2012.2210699

Zhang, H. (2004). *The optimality of naive Bayes*. FLAIRS Conference. www.aaai.org.

Zhang, G., Wang, M., & Liu, K. (2019). Forest fire susceptibility modeling using a convolutional neural network for Yunnan Province of China. *International Journal of Disaster Risk Science, 10*(3), 386–403. https://doi.org/10.1007/s13753-019-00233-1

Zhang, G., Wang, M., & Liu, K. (2021). Deep neural networks for global wildfire susceptibility modelling. *Ecological Indicators, 127*(August), Article 107735. https://doi.org/10.1016/j.ecolind.2021.107735