

# Drivers and distribution of forest fires in Sikkim Himalaya: A maximum entropy-based approach to spatial modelling.

Polash Banerjee (✉ [banerjee.polash@gmail.com](mailto:banerjee.polash@gmail.com))

Sikkim Manipal University

---

## Research Article

**Keywords:** GIS, Likelihood map, Machine learning, MaxEnt, Wildfire

**Posted Date:** October 7th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-22646/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published on April 30th, 2021. See the published version at <https://doi.org/10.1080/02827581.2021.1918239>.

## Abstract

The recent episodes of forest fire in Brazil and Australia of 2019 are tragic reminders of the hazards of the forest fire. Globally incidents of forest fire events are in the rise due to human encroachment into wilderness and climate change. Sikkim with a forest cover of more than 47%, suffers seasonal instances of frequent forest fire during the dry winter months. To address this issue, a GIS-aided and MaxEnt machine learning-based forest fire prediction map has been prepared using forest fire inventory database and maps of environmental features. The study indicates that amongst the environmental features, climatic conditions and proximity to roads are the major determinants of the forest fire. Model validation criteria like ROC curve, correlation coefficient and Cohen's Kappa show a good predictive capability (AUC = 0.95, COR = 0.78,  $\kappa$  = 0.78). The outcomes of this study in the form of a forest fire prediction map can aid the stakeholders of the forest in taking informed mitigation measures.

## 1. Introduction

The incidents of forest fire in Sikkim Himalaya take a peak during the dry period of the year from November to March due to the accumulation of dry biomass over the forest floor. These incidents may occur by natural causes like lightning as Sikkim falls under the northeast region of India, which is considered a high lightening zone. Anthropogenic causes of forest fire in Sikkim include intentional and accidental factors. Bonfires by the cattle herders, burning of the forest floor to deter wild animals entering the agrarian land, logging induced decrease in forest canopy cover are the intentional causes of forest fire in Sikkim. While, sparks from the uphill moving vehicles, electric transformers located in forested areas, use of traditional torch called Rankoo, throwing away of live bidi and cigarettes butts are the accidental causes of forest fire (S. Sharma et al., 2014).

A forest fire can be considered as a mixed blessing. Low-intensity forest fire opens the canopy cover and removes dead wood, providing a niche for new plants to grow. Also, burning of the forest releases the nutrients bound with the biomass to the soil, rejuvenating forest growth. Forest fire also offers new ecological niches for wildlife to proliferate. In contrast, high-intensity forest fire leads to loss of soil biota, volatilization of soil nutrients, increase in soil erosion, a decline in biodiversity and forest biomass (Chandra & Bhardwaj, 2015; Parashar & Biswas, 2003).

A wide range of features have been considered for prediction of forest fire. According to the review of forest fire in the Indian context done by Joseph et al. (2009), topographical features like altitude, aspect, slope, Topographic Wetness Index (TWI) have been used in forest fire prediction. These features are partly accountable for the type and distribution of vegetation. Also, topographic features influence the direction and speed of forest fire (Guo et al., 2016; Maingi & Henry, 2007; Minnich & Bahre, 1995). Meteorological features like average precipitation, temperature, humidity and wind speed have been used to understand forest fire characteristics. In other studies lightening has been focused to predict forest fire (Chen et al., 2015). These meteorological features can act as triggering and conditioning factors for the start and spread of forest fire (Guo et al., 2016; Turco et al., 2013). Vegetational features like vegetation type, Normalized Vegetation Difference Index (NDVI), tree cover fraction; human induced features like proximity to road network, human habitation or Wildland–Urban Interface (WUI); and in-situ factors like soil moisture, soil texture and fuel density have also been used for forest fire prediction (Ghoshlaghi et al., 2020; Jaafari et al., 2018; Mhawej et al., 2015; Satir et al., 2016)

A forest fire or wildfire prediction map has become a valuable tool for disaster management and ecological restoration. Multicriteria decision analysis such as Analytic Hierarchy Process (AHP), Analytical Network Process (ANP) and other forms of expert opinion based methods have been applied in forest fire prediction mapping (Ghoshlaghi et al., 2020; Goleiji et al., 2017; Ljubomir et al., 2019; Regodic et al., 2018; Yathish et al., 2019). In these methods, the model criteria and alternatives are considered as a hierarchical structure. This is followed by the ranking of the model criteria and alternatives based on a certain scale. Based on the ranking the importance or weights of the model criteria and alternatives are estimated and then used in the GIS-aided prediction mapping (Banerjee et al., 2018). However, expert opinion-based prediction mapping may suffer subjective bias. Moreover, these methods are deterministic. As a result, they may not be suitable for a phenomenon that involves uncertainty, such as a forest fire (Ishizaka & Labib, 2009; Mendoza & Martins, 2006). Machine learning methods such as kernel logistic regression, support vector machine, random forest, fuzzy logic, MaxEnt, multilayer perceptron, deep learning and convolutional neural networks have been extensively used in forest fire prediction mapping. Contrary to expert opinion-based methods, machine learning methods do not suffer from subjective bias. Moreover, these methods encompass the uncertainty associated with the modelling of a phenomenon. However, machine learning may suffer issues like model overfitting. These methods heavily rely on the training dataset and take time to learn. Furthermore, these methods require a large dataset of events of interest for proper training of the model. Another important limitation of machine learning method, to be specific methods involving artificial neural networking, is that, they achieve efficiency and accuracy at the cost of interpretability of the model (Nami et al. 2018; Tien Bui, Le, and Hoang 2018; Ghorbanzadeh, Kamran, and Blaschke 2019; Tehrany et al. 2019; Tien Bui, Hoang, and Samui 2019; Zhang, Wang, and Liu 2019).

Maximum entropy or MaxEnt is a popular machine learning method widely being used in species distribution and earth hazard modelling (Feng & Hong, 2009; Harte, 2011; Pourghasemi & Rossi, 2018). Unlike most machine learning methods such as logistic regression, support vector machine, random forest, k-nearest neighbour and artificial neural network, that uses presence-absence instances dataset for training, the MaxEnt uses presence-background instances dataset for training. MaxEnt is based on the principle, that the probability distribution that maximizes entropy for the current state of knowledge subject to the constraints of the features is the best fit model for the phenomenon under consideration (De Martino & De Martino, 2018). It is popular primarily because it considers 'minimum assumption' while selecting a probability distribution (Warton, 2013). Moreover, this method considers more realistic presence-background dataset, in the sense that in nature hardly any absence data is available. On the other hand, MaxEnt needs a large presence-dataset to perform reliable prediction. Also, a study has suggested that MaxEnt is equivalent to the Generalized Linear Model (GLM) when it comes to Point Process Models (PPMs) such as forest fire events (Fithian & Hastie, 2013). MaxEnt has been used in several forest fire prediction mappings. Studies indicate that MaxEnt has performed equally well in comparison to other machine learning methods in predicting a forest fire. (Arpacı et al., 2014; Fernández-Manso & Quintano, 2020; Fonseca et al., 2016; Kim et al., 2015; Lim et al., 2017; Massada et al., 2013; Peters et al., 2013)

In this study, MaxEnt has been applied to prepare a forest fire prediction map of Sikkim Himalaya using MODIS and Ground data-based forest fire inventory. As features, meteorological, topological, ecological, in-situ and human-induced data have been used to train the MaxEnt model. Model validation criteria have been used to evaluate the model. The study indicates that MaxEnt is a reliable machine learning method in predicting areas prone to forest fire events in Sikkim Himalaya.

## 2. Materials And Methods

- Study area

Sikkim is a small eastern Himalayan state of India neighboured by Tibet in the North, Nepal in the West, Bhutan in the East and the state of West Bengal in the south. It extends from 27° 00' 46" N to 28° 07' 48" N latitude and 88° 00' 58" E to 88° 55' 25" E longitude. The elevation of Sikkim varies from 280 m in the South to 8586 m in the North, crowned by the world's third-highest mountain peak, Mt. Khangchendzonga (Shukla et al., 2018). Sikkim, apart from having four seasons of winter, summer, spring, autumn, has a monsoon season lasting from June to September. It has a sub-tropical climate in the south and tundra climate in the north. The two main rivers of Sikkim include the Teesta River and its tributary, the Rangeet (ENVIS Sikkim, 2019) (**Figure 1a-b**).

Sikkim is endowed with various vegetation ecotypes based on the elevation and climatic conditions, like Himalayan subtropical broadleaf forests in the lower elevations, Eastern Himalayan broadleaf forests in the temperate zone above the elevation of 1500 metres, Eastern Himalayan subalpine conifer forests from 3500 to 5000 metres and Eastern Himalayan alpine shrub and meadows in the higher elevations (O'Neill, 2019; O'Neill et al., 2020). The dry winter season from December to March in Sikkim characterised by windy weather and dry forest biomass, create the right conditions for a forest fire. It is more common in the deciduous Sal forest ecosystem followed by the temperate oak and sub-alpine conifer forests. Erratic rainfall, climate change and conversion of forest land into other land uses have increased the vulnerability of the forests in Sikkim, leading to a growing trend of forest fire incidents (**Figure 2**) (Banerjee et al., 2020; R. Sharma et al., 2012).

- Data sources

The active fire data from the year 2000-2019 of Moderate Resolution Imaging Spectroradiometer (MODIS) was accessed from the data archive at the Fire Information for Resource Management System (FIRMS) site. The MODIS dataset was combined with the forest fire inventory prepared from the GPS-tagged dataset of the Forest and Environment Department, Government of Sikkim. The fire incidents dataset thus prepared was intersected with the forest fraction raster data (Shimada et al., 2014) to exclude fire incidents beyond the forest cover of the study area. This generated a fire dataset of 754 events.

The environmental feature raster maps or simply features used in this study included precipitation (*avgPrep*), ambient air temperature (*avgTemp*) and wind speed (*avgWind*) averaged over the dry period of Sikkim prepared from the monthly data accessed from Worldclim 2 (Fick & Hijmans, 2017). For this study, amongst several data resolutions, the 30 seconds resolution average monthly climate data for 1970-2000 was taken from Worldclim 2. Also, features like aspect, plan curvature (*PlanCurv*), profile curvature (*profileCurv*), slope and TWI were derived from the Digital Elevation Model (DEM) (Jarvis, A., H.I. Reuter, A., Nelson, E. Guevara. 2008). The DEM used in this study was the product of Shuttle Radar Topography Mission (SRTM) of 90m resolution derived from CGIAR/SRTM90\_V4 image with a data collection timeframe from 2000-02-11 to 2000-02-22. The NDVI of the study area was prepared from the Advanced Spaceborne Thermal Emission and Reflection Radiometer data (ASTER Mount Gariwang image, 2018). The NDVI data of 250m resolution was prepared from MODIS/006/MOD13Q1 image, computed from atmospherically corrected bi-directional surface reflectance that has been masked for water, clouds, heavy aerosols, and cloud shadows. The timeframe of NDVI data collection was from 2000-02-18 to 2020-07-02. Features like percent tree cover (Sexton et al., 2013) and population density (CIESIN 2018) were also used in the modelling. The percent tree cover (*treeCov*) was the product of 30m resolution image of GLCF/GLS\_TCC of Global Landsat Tree Cover Continuous Fields, prepared from three-year epochs of 2000, 2005 and 2010 assessing the woody vegetation greater than 5 meters in height. In-situ features, namely, soil surface carbon content (*soilCarbon*) (Tomislav Hengl & Ichsanji Wheeler, 2018) and soil surface water content (*soilWater*) (Tomislav Hengl & Surya Gupta, 2019) were accessed from OpenLandMap. The population density data was availed from the Gridded Population of the World, Version 4.11 (GPWv4) released on December 2018 prepared by the Center for International Earth Science Information Network (CIESIN), Columbia University (CIESIN, 2018). Features like vector datasets of the road network, waterways and human habitations were accessed from © OpenStreetMap contributors and used to prepare proximity raster maps as *proxRoads*, *proxWater* and *proxPlace* of the respective features (**Figure 4 a-n**).

- Maximum entropy – the MaxEnt Model

The MaxEnt algorithm does prediction by minimizing the relative entropy between the probability density of the presence-only instances of the target variable from that of the instances of background landscape data (Elith et al., 2011). For instances, for a landscape,  $L$  the algorithm uses forest fire occurrence, ( $y = 1$ ) over a vector of environmental features,  $\mathbf{z}$ . The MaxEnt algorithm attempts to minimize the distance of the probability densities of features in case of forest fire occurrences, from the probability densities of features of the background or the null model, over the landscape  $L$ . The minimization of the distance function is achieved by maximizing a penalized likelihood model subject to the model constraints given by the probability densities of features of the landscape (Steven J. Phillips & Dudík, 2008).

Unlike conventional machine learning methods like logistic regression or random forest which uses the presence-absence data, MaxEnt uses presence-background data for prediction of the forest fire. This makes MaxEnt prone to sample selection bias, a condition where some areas of the landscape may be over-sampled than other areas. Also, MaxEnt is prone to overfitting the predictive model to the presence-only data (Devisscher et al., 2016). However, recent MaxEnt software uses transformation algorithms like linear, quadratic, product, categorical, threshold and hinge for the standardization of the prediction features and regularization of the model to prevent model overfitting (Elith et al., 2011).

Forest fires, like most natural phenomena, rarely have data on their absence. This makes MaxEnt more appropriate for forest fire studies as it requires presence-only data for their predictions (Arnold et al., 2014; Steven J. Phillips & Elith, 2013).

- **Data processing and preparation of forest fire prediction map**

Initially, all feature maps were projected from geographic projection system of GCS-WGS-1984 to plane projection system of WGS-1984-UTM-Zone-45N, which is suitable for the study area. Next, Euclidean distance raster maps were prepared from the polyline vector maps of the road network and waterbodies network and point vector map of human habitations. These raster maps were prepared to measure the proximity of fire events from roads, waterbodies and human habitations. Topographic feature maps of aspect, slope and TWI were prepared from DEM. Thereafter, all the feature maps were changed to have the same cell size and same extent. Next, the feature maps were normalised, such that the pixel values of the maps were in the range from zero to one (Chang, 2017). The normalized maps were exported in GeoTiff format as they are readily readable by R-programming language. Furthermore, the presence-only dataset of forest fire events was stored as a CSV file.

The forest fire prediction map was prepared in RStudio environment using R packages named 'raster' (Hijmans, 2020), 'rgdal' (Bivand et al., 2019) and 'dismo' (Hijmans et al., 2017). The first two packages were mainly used for raster images related spatial operations while dismo was used for bridging between R and MaxEnt software. The MaxEnt software used in this study is a java program-based package (S.J. Phillips et al., n.d.). During the preparation of the prediction map, all the feature maps were stacked with matching extents and feature attributes were extracted from the features stack using the fire event coordinates. The fire dataset was divided into five-folds for crossvalidation. This was followed by the preparation of background dataset by selecting 1000 random points from the extent of the study area. Similar to the preparation of fire event dataset, the background dataset was populated with the feature attributes and divided into five-fold datasets for crossvalidation. In this process, repetitively any one sub-dataset out of the five sub-dataset is used for testing while the rest are used for training the MaxEnt algorithm. An average of the errors generated from the repeated tests help in the tuning of the model parameters for better performance of MaxEnt.

The MaxEnt method-based prediction was applied to the entire extent of the study area, considering every cell as an instant. The model output was exported as a GeoTiff raster and classified into five qualitative categories applying natural breaks method of ArcGIS framework. All Model validation criteria like Receiver Operating Characteristic (ROC) curve (Peres & Cancelliere, 2014), correlation coefficient and Cohen's kappa (Vakhshoori & Zare, 2018) were used to evaluate the model performance. ROC curve is a popular diagnostic and visualization tool that plots sensitivity of the model against specificity of the model. The correlation coefficient compares how well the predicted values match with the observed values. Cohen's Kappa statistic measures the agreement between two categorical scales, such as binary outcomes of predicted and observed events of forest fire. The kappa is the ratio of the deviation of the predicted value from the observed to one less predicted value (Sim & Wright, 2005). In all the three cases, a value close to one is satisfactory for model validation.

Moreover, importance of the features of the model and sensitivity analysis were performed. The methodology of the study is illustrated below (**Figure 3**).

### 3. Results

Correlation analysis indicated that feature like popDen was strongly correlated with avgPrep and avgTemp, primarily because most of the human population was in the south of Sikkim. Also, avgTemp had a strong correlation with avgPrep, as the subtropical Sikkim gets the bulk of rainfall (**Table 1**). Multicollinearity analysis indicated that Variance Inflation Factor (VIF) of the independent feature variables against avgTemp was within acceptable limits ( $VIF < 5$ ), except for popDen. Hence, multicollinear variable like population density was dropped from the model (**Table 2**).

Starting with the meteorological features, the fire events were more common in moderate to warmer areas ( $11 - 24^{\circ}\text{C}$ ). Like avgTemp, the fire events were more common in areas with moderate to higher avgPrep (35 - 55mm). In contrast, fire events were common in areas with low avgWind ( $1.4 - 1.7 \text{ ms}^{-1}$ ). Moving onto topographic features, bulk of the fire events were observed in the flatter slope (5 - 7 degree), lower TWI (4 - 6 value) and moderate aspect (81 - 217 degree). The PlanCurv and profileCurv had nominal influence in the prediction of forest fires. In case of PlanCurv, convex curvature of 0.765 had some influence on forest fire occurrences. Concave curvature of 0.488 of profileCurv had some influence in predicting forest fires. Looking at the ecological features, fire events were common to areas with moderate to high NDVI value (0.5 - 0.7) having moderate treeCov (31 - 55% of the area). In-site features indicated that areas with moderate soilCarbon and low soilWater were more common to have forest fires. Furthermore, fire events were skewed towards areas close to the waterbodies (0 - 800m), human habitations (0 - 3000m) and roadways (0 - 800m) (**Figure 5, supplement**).

In terms of contribution of the features towards prediction of forest fire, proxRoads explained 43% of events, followed by avgTemp that explained almost 17% of fire events. Environmental features, namely, proxRoads, avgTemp, treeCov, avgPrep, proxPlace and avgWind together explained 85% of the forest fire events. TWI had no contribution in the prediction value. Hence, TWI was dropped from the MaxEnt model (**Figure 6**).

The MaxEnt method-based forest fire prediction map of 30.7m resolution showed a probability range from 0 to 1, indicating no chances of forest fire occurrences to very high chances of forest fire occurrences (**Figure 7**). The prediction map was further categorized into very low, low, medium, high and very high chances of forest fire incidents for the sake of convenience (**Figure 8**). Out of the total study area,  $351 \text{ km}^2$  was under very high probability of forest fire, while  $451 \text{ km}^2$  was under high probability of forest fire. Thereby, these two categories together constitute 11% of the total study area.

Model validation criteria like ROC curve showed an Area Under the Curve (AUC) of 0.957 (**Figure 9a**). The correlation coefficient was estimated to be 0.809 and Cohen's Kappa was estimated to be 0.78 (**Figure 9b**).

### 4. Discussion

In this study an attempt has been made to prepare the forest fire prediction map of Sikkim Himalaya using MaxEnt machine learning method. The study indicated that estimation of probability of forest fire by MaxEnt was satisfactory as per the model validation criteria.

It was observed that out of the 15 predictor features, six of them, namely proxRoad, avgTemp, treeCov, avgPrep, proxPlace, avgWind, were able to explain 85% of forest fire events. Due to the difficult terrain, much of the human activities in Sikkim are confined around the roadways. This also includes the human habitations. The limited contribution of topographic and ecological features in explaining the forest fires further reinforces the disproportionate influence of human activities and meteorological factors in the occurrences of forest fire. Together, these two anthropogenic features explain 47.5% of forest fires. This indicates with a fair amount of confidence, that forest fires in Sikkim Himalaya primarily occur due to human activities and they are facilitated by climatic conditions. This observation was similar to a previous study done in the Amazonian forest of Bolivia (Devisscher et al., 2016). Also, studies conducted in the Huron-Manistee National Forest, Michigan, USA suggested that development activities such as road networks were the major determinants of forest fire occurrences (Massada et al., 2013). Other studies on forest fire prediction modelling also suggest that proximity to roads and human settlements are the major determinants of forest fires (Arpacı et al., 2014; Jaafari et al., 2018; Maingi & Henry, 2007). Areas with low to moderate tree cover were more susceptible to forest fire. Also, areas with moderate temperature and precipitation with very low wind speed were more prone to forest fires. Again, these areas are prevalent in the valleys of southern Sikkim Himalaya which are dominated by agrarian land, human habitations and road networks. This observation further indicates that the ecological and meteorological conditions are the conditioning factors of forest fire in Sikkim.

In contrast to the mainland India where forest fire is common during the hot dry summer (Joseph et al., 2009), majority of forest fire in Sikkim Himalaya occurs during the cold dry period from November to March. As observed in this study, bulk of the forest fire events were in the southern part of Sikkim. This is mainly due to the logging activities there. Also, higher vehicular traffic explained by greater road network in southern Sikkim makes the dry vegetation vulnerable to fire due to engine spark and cigarette or bidi butts. The limited number of forest fire events in the high altitude of Sikkim is primarily due to lightning. Moreover, the high contrast of warmer climate in southern Sikkim as compared to very cold climatic conditions of northern Sikkim makes the former more vulnerable to forest fire (R. Sharma et al., 2012). This study also indicated that forested areas close to human habitations are at a higher risk of forest fire. An aspect from East to South-West direction had more contribution towards forest fire. Aspect influences soil moisture, solar radiation, vegetation composition and density (Estes et al., 2017). Also, forest patches of valley areas that receive moderate rainfall, have moderate temperature and low wind speed were prone to forest fire. The higher values of model validation criteria suggested that the model prediction was satisfactory.

Being fundamentally distinct from other machine learning methods, MaxEnt uses presence-only dataset to train itself (Elith et al., 2011). However, like many studies have shown earlier, the present study also indicated that this distinction of MaxEnt does not limit its capability in generating reliable hazard prediction maps (Arpacı et al., 2014; Fernández-Manso & Quintano, 2020; Fonseca et al., 2016; Kim et al., 2015; Lim et al., 2017; Massada et al., 2013; Peters et al., 2013). In this study a limited set of features have been considered for forest fire prediction. This was primarily due to availability of reliable data. However, other features like lightning activities in the North Sikkim, dry fuel biomass and vegetation type could be considered to improve the model.

The forest fire prediction map of Sikkim Himalaya can be considered as a decision support tool for stakeholders of forest resources. The forest managers such as forest rangers and forest dependent communities can mitigate forest fire by allocating their fire control resources to areas more prone to forest fire. Population of forest fire prone areas can be educated about the impacts of their activities on the occurrence of forest fire. Targeted law enforcement against irresponsible activities like illegal logging, negligent smoking and bonfire, slash and burn farming and traffic management can be achieved from the forest fire prediction map.

## 5. Conclusion

Applications of remote sensing imageries, machine learning and geospatial analysis can mitigate forest fire by identifying areas that were relatively more prone to forest fire. MaxEnt-based forest fire prediction map of Sikkim Himalaya indicated that anthropogenic features, mainly road network, tree cover fraction and meteorological features were mainly accountable for forest fire incidents. Although a forest fire can be considered as an opportunity for the forest to rejuvenate, increase in frequency and extent of forest fire can only lead to damage to the forest health. The prediction map can be used as a decision support tool by the stakeholders to mitigate the occurrences forest fire. The applications of MaxEnt can be extended to other forms of earth hazards like landslide, flood and drought predictions. The MaxEnt model can be further improved by expanding the feature set, followed by factor analysis to identify the most relevant explanatory features of forest fire. A comparative analysis of the MaxEnt along with other machine learning methods can be performed to assess the efficiency and efficacy of MaxEnt. The outcomes of this study can be internalized into forest management policies by applying geographically targeted resource allocation and law enforcement towards forest fire mitigation.

## References

- Arnold, J. D., Brewer, S. C., & Dennison, P. E. (2014). Modeling Climate-Fire Connections within the Great Basin and Upper Colorado River Basin, Western United States. *Fire Ecology*, 10(2), 64–75. <https://doi.org/10.4996/fireecology.1002064>
- Arpacı, A., Malowerschnig, B., Sass, O., & Vacik, H. (2014). Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. *Applied Geography*, 53, 258–270. <https://doi.org/10.1016/j.apgeog.2014.05.015>
- ASTER Mount Gariwang image. (2018). MOD13Q1.006 Terra Vegetation Indices 16-Day Global 250m; NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC). USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. <https://doi.org/10.5067/MODIS/MOD13Q1.006>

- Banerjee, P., Ghose, M. K., & Pradhan, R. (2018). Analytic hierarchy process and information value method-based landslide susceptibility mapping and vehicle vulnerability assessment along a highway in Sikkim Himalaya. *Arabian Journal of Geosciences*, 11(7), 139. <https://doi.org/10.1007/s12517-018-3488-4>
- Banerjee, P., Ghose, M. K., & Pradhan, R. (2020). Analytic hierarchy process based spatial biodiversity impact assessment model of highway broadening in Sikkim Himalaya. *Geocarto International*, 35(5), 470–493. <https://doi.org/10.1080/10106049.2018.1520924>
- Bivand, R., Keitt, T., & Rowlingson, B. (2019). rgdal: Bindings for the “Geospatial” Data Abstraction Library. <https://CRAN.R-project.org/package=rgdal>
- Chandra, K. K., & Bhardwaj, A. K. (2015). Incidence of Forest Fire in India and Its Effect on Terrestrial Ecosystem Dynamics, Nutrient and Microbial Status of Soil. *International Journal of Agriculture and Forestry*, 5(2), 69–78.
- Chang, K.-T. (2017). Introduction to Geographic Information Systems (4 edition). McGraw Hill Education.
- Chen, F., Du, Y., Niu, S., & Zhao, J. (2015). Modeling Forest Lightning Fire Occurrence in the Daxinganling Mountains of Northeastern China with MAXENT. *Forests*, 6(5), 1422–1438. <https://doi.org/10.3390/f6051422>
- CIESIN. (2018). Gridded Population of the World, Version 4 (GPWv4): Basic Characteristics, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Columbia University Center for International Earth Science Information Network (CIESIN) - Columbia University. <https://doi.org/10.7927/H46M34XX>
- De Martino, A., & De Martino, D. (2018). An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon*, 4(4), e00596. <https://doi.org/10.1016/j.heliyon.2018.e00596>
- Devisscher, T., Anderson, L. O., Aragão, L. E. O. C., Galván, L., & Malhi, Y. (2016). Increased Wildfire Risk Driven by Climate and Development Interactions in the Bolivian Chiquitania, Southern Amazonia. *PLOS ONE*, 11(9), e0161323. <https://doi.org/10.1371/journal.pone.0161323>
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>
- ENVIS Sikkim. (2019). Rivers of Sikkim [Map]. [http://sikenvis.nic.in/Database/Rivers\\_781.aspx](http://sikenvis.nic.in/Database/Rivers_781.aspx)
- 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), e01794. <https://doi.org/10.1002/ecs2.1794>
- Feng, L., & Hong, W. (2009). On the principle of maximum entropy and the risk analysis of disaster loss. *Applied Mathematical Modelling*, 33(7), 2934–2938. <https://doi.org/10.1016/j.apm.2008.10.002>
- Fernández-Manso, A., & Quintano, C. (2020). A Synergetic Approach to Burned Area Mapping Using Maximum Entropy Modeling Trained with Hyperspectral Data and VIIRS Hotspots. *Remote Sensing*, 12(5), 858. <https://doi.org/10.3390/rs12050858>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- Fithian, W., & Hastie, T. (2013). Finite-sample equivalence in statistical models for presence-only data. *The Annals of Applied Statistics*, 7(4), 1917–1939. <https://doi.org/10.1214/13-AOAS667>
- Fonseca, M. G., Aragão, L. E. O. C., Lima, A., Shimabukuro, Y. E., Arai, E., & Anderson, L. O. (2016). Modelling fire probability in the Brazilian Amazon using the maximum entropy method. *International Journal of Wildland Fire*, 25(9), 955–969. <https://doi.org/10.1071/WF15216>
- Gheshlaghi, H. A., Feizizadeh, B., & Blaschke, T. (2020). GIS-based forest fire risk mapping using the analytical network process and fuzzy logic. *Journal of Environmental Planning and Management*, 63(3), 481–499. <https://doi.org/10.1080/09640568.2019.1594726>
- Ghorbanzadeh, O., Kamran, K. V., & Blaschke, T. (2019). Spatial Prediction of Wildfire Susceptibility Using Global NASA MODIS Fire Products and Machine Learning Approaches. <https://uni-salzburg.elsevierpure.com/en/publications/spatial-prediction-of-wildfire-susceptibility-using-global-nasa-m>
- Goleiji, E., Hosseini, S. M., Khorasani, N., & Monavari, S. M. (2017). Forest fire risk assessment-an integrated approach based on multicriteria evaluation. *Environmental Monitoring and Assessment*, 189(12), 612. <https://doi.org/10.1007/s10661-017-6225-7>
- Guo, F., Wang, G., Su, Z., Liang, H., Wang, W., Lin, F., & Liu, A. (2016). What drives forest fire in Fujian, China? Evidence from logistic regression and Random Forests. *International Journal of Wildland Fire*, 25(5), 505–519. <https://doi.org/10.1071/WF15121>
- Harte, J. (2011). Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics (1 edition). Oxford University Press.
- Hijmans, R. J. (2020). raster: Geographic Data Analysis and Modeling. <https://CRAN.R-project.org/package=raster>
- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). dismo: Species Distribution Modeling. <https://CRAN.R-project.org/package=dismo>

- Ishizaka, A., & Labib, A. (2009). Analytic Hierarchy Process and Expert Choice: Benefits and limitations. *OR Insight*, 22(4), 201–220. <https://doi.org/10.1057/ori.2009.10>
- Jaafari, A., Zenner, E. K., & Pham, B. T. (2018). Wildfire spatial pattern analysis in the Zagros Mountains, Iran: A comparative study of decision tree based classifiers. *Ecological Informatics*, 43, 200–211. <https://doi.org/10.1016/j.ecoinf.2017.12.006>
- Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara. (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database. <http://srtm.csi.cgiar.org>.
- Joseph, S., Anitha, K., & Murthy, M. S. R. (2009). Forest fire in India: A review of the knowledge base. *Journal of Forest Research*, 14(3), 127–134. <https://doi.org/10.1007/s10310-009-0116-x>
- Kim, T., Lim, C. H., Song, C., & Lee, W. K. (2015). Estimation of Wild Fire Risk Area based on Climate and Maximum Entropy in Korean Peninsular. *AGU Fall Meeting Abstracts*, 31, NH31A-1880.
- Lim, C. H., Kim, M., Kim, S. J., Yoo, S., & Lee, W. K. (2017). Assessment of multi-wildfire occurrence data for machine learning based risk modelling. *AGU Fall Meeting Abstracts*, 42. <http://adsabs.harvard.edu/abs/2017AGUFMNH42B..10L>
- Ljubomir, G., Pamučar, D., Drobnjak, S., & Pourghasemi, H. R. (2019). 15—Modeling the Spatial Variability of Forest Fire Susceptibility Using Geographical Information Systems and the Analytical Hierarchy Process. In H. R. Pourghasemi & C. Gokceoglu (Eds.), *Spatial Modeling in GIS and R for Earth and Environmental Sciences* (pp. 337–369). Elsevier. <https://doi.org/10.1016/B978-0-12-815226-3.00015-6>
- Maingi, J. K., & Henry, M. C. (2007). Factors influencing wildfire occurrence and distribution in eastern Kentucky, USA. *International Journal of Wildland Fire*, 16(1), 23–33. <https://doi.org/10.1071/WF06007>
- Massada, A. B., Syphard, A. D., Stewart, S. I., & Radeloff, V. C. (2013). Wildfire ignition-distribution modelling: A comparative study in the Huron–Manistee National Forest, Michigan, USA. *International Journal of Wildland Fire*, 22(2), 174–183. <https://doi.org/10.1071/WF11178>
- Mendoza, G. A., & Martins, H. (2006). Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest Ecology and Management*, 230(1), 1–22. <https://doi.org/10.1016/j.foreco.2006.03.023>
- Mhawej, M., Faour, G., & Adjizian-Gerard, J. (2015). Wildfire Likelihood's Elements: A Literature Review. *Challenges*, 6(2), 282–293. <https://doi.org/10.3390/challe6020282>
- Minnich, R. A., & Bahre, C. J. (1995). Wildland Fire and Chaparral Succession Along the California Baja-California Boundary. *International Journal of Wildland Fire*, 5(1), 13–24. <https://doi.org/10.1071/wf9950013>
- Nami, M. H., Jaafari, A., Fallah, M., & Nabiuni, S. (2018). Spatial prediction of wildfire probability in the Hycanian ecoregion using evidential belief function model and GIS. *International Journal of Environmental Science and Technology*, 15(2), 373–384. <https://doi.org/10.1007/s13762-017-1371-6>
- O'Neill, A. R. (2019). Evaluating high-altitude Ramsar wetlands in the Eastern Himalayas. *Global Ecology and Conservation*, 20, e00715. <https://doi.org/10.1016/j.gecco.2019.e00715>
- O'Neill, A. R., Chhetri, P. K., Chhetri, B., & Rana, S. K. (2020). Establishing ecological baselines around a temperate Himalayan peatland. *Wetlands Ecology and Management*. <https://doi.org/10.1007/s11273-020-09710-7>
- Parashar, A., & Biswas, S. (2003). The Impact Of Forest Fire On Forest Biodiversity In The Indian Himalayas (Uttaranchal). *Proceeding of XII World Forestry Congress*, Canada 0358-B1. <http://www.fao.org/3/XII/0358-B1.htm>
- Peres, D. J., & Cancelliere, A. (2014). Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach. *Hydrology and Earth System Sciences*, 18(12), 4913–4931. <https://doi.org/10.5194/hess-18-4913-2014>
- Peters, M. P., Iverson, L. R., Matthews, S. N., & Prasad, A. M. (2013). Wildfire hazard mapping: Exploring site conditions in eastern US wildland–urban interfaces. *International Journal of Wildland Fire*, 22(5), 567–578. <https://doi.org/10.1071/WF12177>
- Phillips, S.J., Dudík, M., & Schapire, R. E. (n.d.). [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Retrieved February 7, 2020, from [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/)
- Phillips, Steven J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, 31(2), 161–175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>
- Phillips, Steven J., & Elith, J. (2013). On estimating probability of presence from use—Availability or presence—Background data. *Ecology*, 94(6), 1409–1419. JSTOR.
- Pourghasemi, H. R., & Rossi, M. (2018). *Natural Hazards GIS-Based Spatial Modeling Using Data Mining Techniques*. Springer.

Regodic, L., Gigovic, G., Jakovljevic, D., & Sekulovic, M. (2018). GIS Multi-Criteria Analysis for Identifying and Mapping Forest Fire Hazard: Nevesinje, Bosnia and Herzegovina. *Tehnicki Vjesnik - Technical Gazette*, 25(3), 891–898.

Satir, O., Berberoglu, S., & Donmez, C. (2016). Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem. *Geomatics, Natural Hazards and Risk*, 7(5), 1645–1658. <https://doi.org/10.1080/19475705.2015.1084541>

Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K. M., Channan, S., DiMiceli, C., & Townshend, J. R. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6(5), 427–448. <https://doi.org/10.1080/17538947.2013.786146>

Sharma, R., Sharma, N., Shrestha, D., Luitel, K., Arrawatia, M., & Pradhan, S. (2012). Study of Forest fires in Sikkim Himalayas, India using Remote sensing and GIS techniques (pp. 233–244).

Sharma, S., Joshi, V., & Chhetri, R. (2014). Forest fire as a potential environmental threat in recent years in Sikkim, Eastern Himalayas, India. *Climate Change and Environmental Sustainability*, 2, 55. <https://doi.org/10.5958/j.2320-642X.2.1.006>

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., & Lucas, R. (2014). New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment*, 155, 13–31. <https://doi.org/10.1016/j.rse.2014.04.014>

Shukla, A., Garg, P. K., & Srivastava, S. (2018). Evolution of Glacial and High-Altitude Lakes in the Sikkim, Eastern Himalaya Over the Past Four Decades (1975–2017). *Frontiers in Environmental Science*, 6. <https://doi.org/10.3389/fenvs.2018.00081>

Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3), 257–268. <https://doi.org/10.1093/ptj/85.3.257>

Tehrany, M. S., Jones, S., Shabani, F., Martínez-Álvarez, F., & Tien Bui, D. (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. *Theoretical and Applied Climatology*, 137(1), 637–653. <https://doi.org/10.1007/s00704-018-2628-9>

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). *Journal of Environmental Management*, 237, 476–487. <https://doi.org/10.1016/j.jenvman.2019.01.108>

Tien Bui, D., Le, H. V., & Hoang, N.-D. (2018). GIS-based spatial prediction of tropical forest fire danger using a new hybrid machine learning method. *Ecological Informatics*, 48, 104–116. <https://doi.org/10.1016/j.ecoinf.2018.08.008>

Tomislav Hengl, & Ichsan Wheeler. (2018). Soil organic carbon content in x 5 g / kg at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.2525553>

Tomislav Hengl, & Surya Gupta. (2019). Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.2784001>

Turco, M., Llasat, M. C., von Hardenberg, J., & Provenzale, A. (2013). Impact of climate variability on summer fires in a Mediterranean environment (northeastern Iberian Peninsula). *Climatic Change*, 116(3), 665–678. <https://doi.org/10.1007/s10584-012-0505-6>

Vakhshoori, V., & Zare, M. (2018). Is the ROC curve a reliable tool to compare the validity of landslide susceptibility maps? *Geomatics, Natural Hazards and Risk*, 9(1), 249–266. <https://doi.org/10.1080/19475705.2018.1424043>

Warton, D. (2013). Some big news about MAXENT. *Methods.Blog*. <https://methodsblog.com/2013/02/20/some-big-news-about-maxent/>

Yathish, H., Athira, K. V., Preethi, K., Pruthviraj, U., & Shetty, A. (2019). A Comparative Analysis of Forest Fire Risk Zone Mapping Methods with Expert Knowledge. *Journal of the Indian Society of Remote Sensing*, 47(12), 2047–2060. <https://doi.org/10.1007/s12524-019-01047-w>

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>

## Tables

Table 1: cross-correlation analysis of the environmental feature.

aspect	1.000												
avgPrep	-0.006												
avgTemp	-0.010												
avgWind	-0.016												
NDVI	0.075												
PlanCurv	0.008												
popDen	0.045												
profileCurv	0.003												
proxPlace	0.013												
proxRoads	-0.056												
proxWater	0.008												
slope	-0.097												
soilCarbon	-0.041												
soilWater	-0.042												
treeCov	0.144												
TWI	0.042												
aspect	avgPrep	avgTemp	avgWind	NDVI	PlanCurv	popDen	profileCurv	proxPlace	proxRoads	proxWater	slope	soilCarbon	soil

Table 2: Multicollinearity analysis of environmental features.

Features	VIF*
aspect	1.054
avgPrep	1.742
avgWind	2.416
NDVI	2.815
PlanCurv	1.612
profileCur	1.181
proxPlace	1.904
proxRoads	1.943
proxWater	1.249
slope	1.388
soilCarbon	1.651
soilWater	1.071
treeCov	1.926
TWI	1.731

\*Independent variable is avgTemp

## Declarations

Ethics approval and consent to participate: Due permission has been taken from the competition authority.

Consent for publication: Due permission has been taken from the university.

Availability of data and material:

<https://drive.google.com/open?id=1jc4AN4ZPClcZ150VQrR1szrEFeGbrSNB>

Competing interests: None

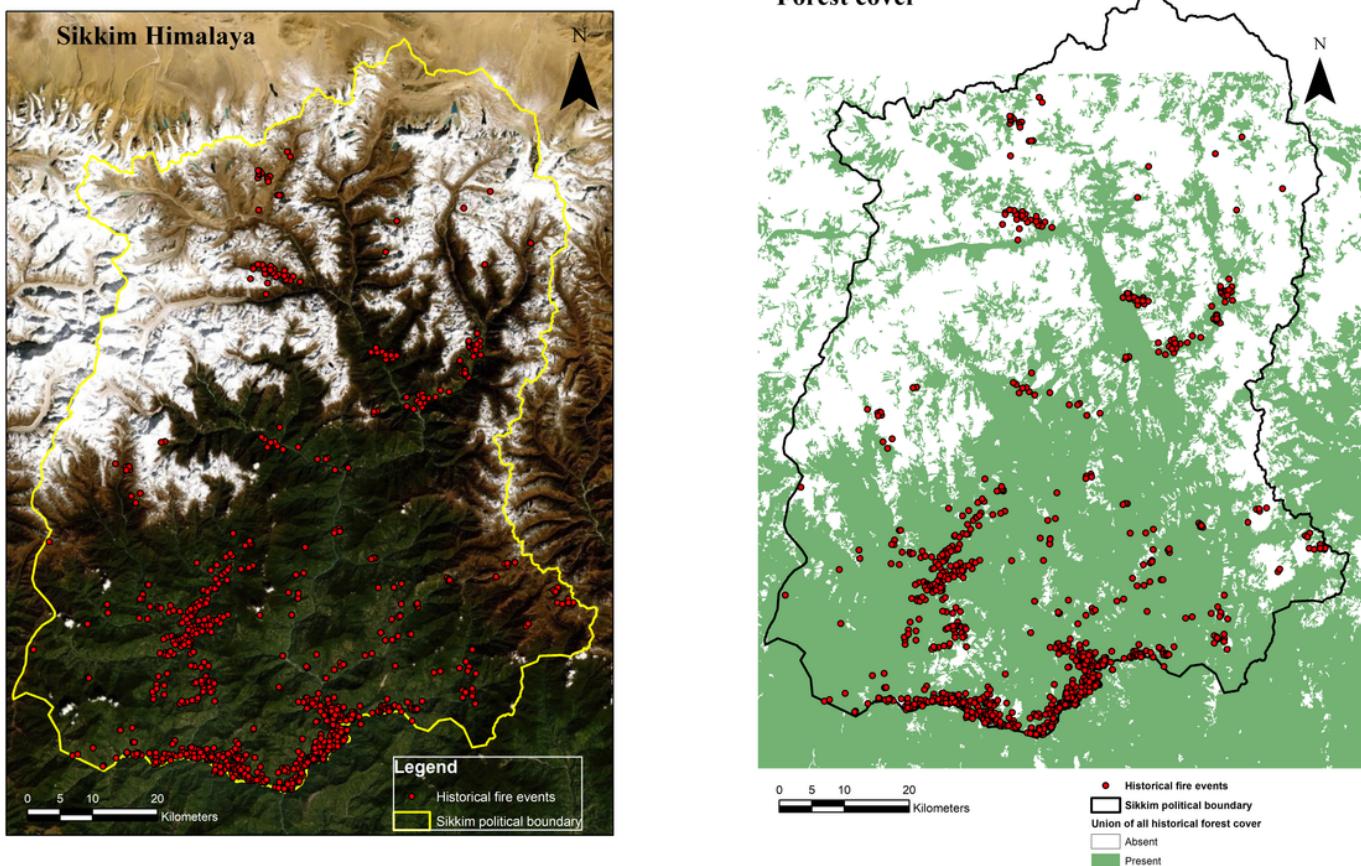
Funding: None

Authors' contributions: Planning, data gathering, data processing, modelling and interpretation.

Acknowledgements:

I acknowledge the use of data and imagery from LANCE FIRMS operated by NASA's Earth Science Data and Information System (ESDIS) with funding provided by NASA Headquarters.

## Figures



**Figure 1**

Study area. (a) Sikkim Himalaya is crowned in the north by snow covered mountains including Mt. Khangchendzonga. The southern part of Sikkim has a wide variety of vegetation including several endemic flora and fauna, and waterbodies in the form of streams and rivers. Human presence is reflected in the form of agrarian land, human habitations, dams, and a dense network of roadways. The forest fire events take certain pattern clustered along the roadways and valley areas. (b) Forest fire events over the historical distribution of forest cover.

Time series and forecast of forest fire events in Sikkim Himalaya

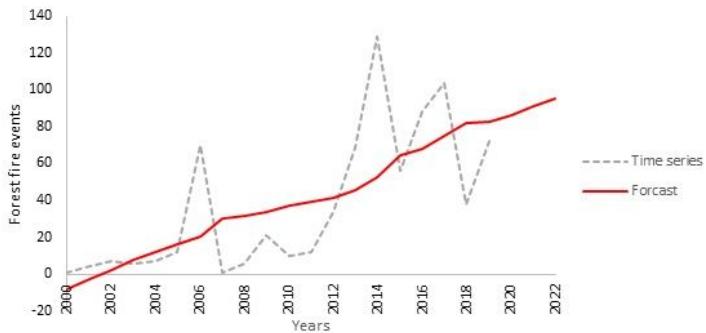
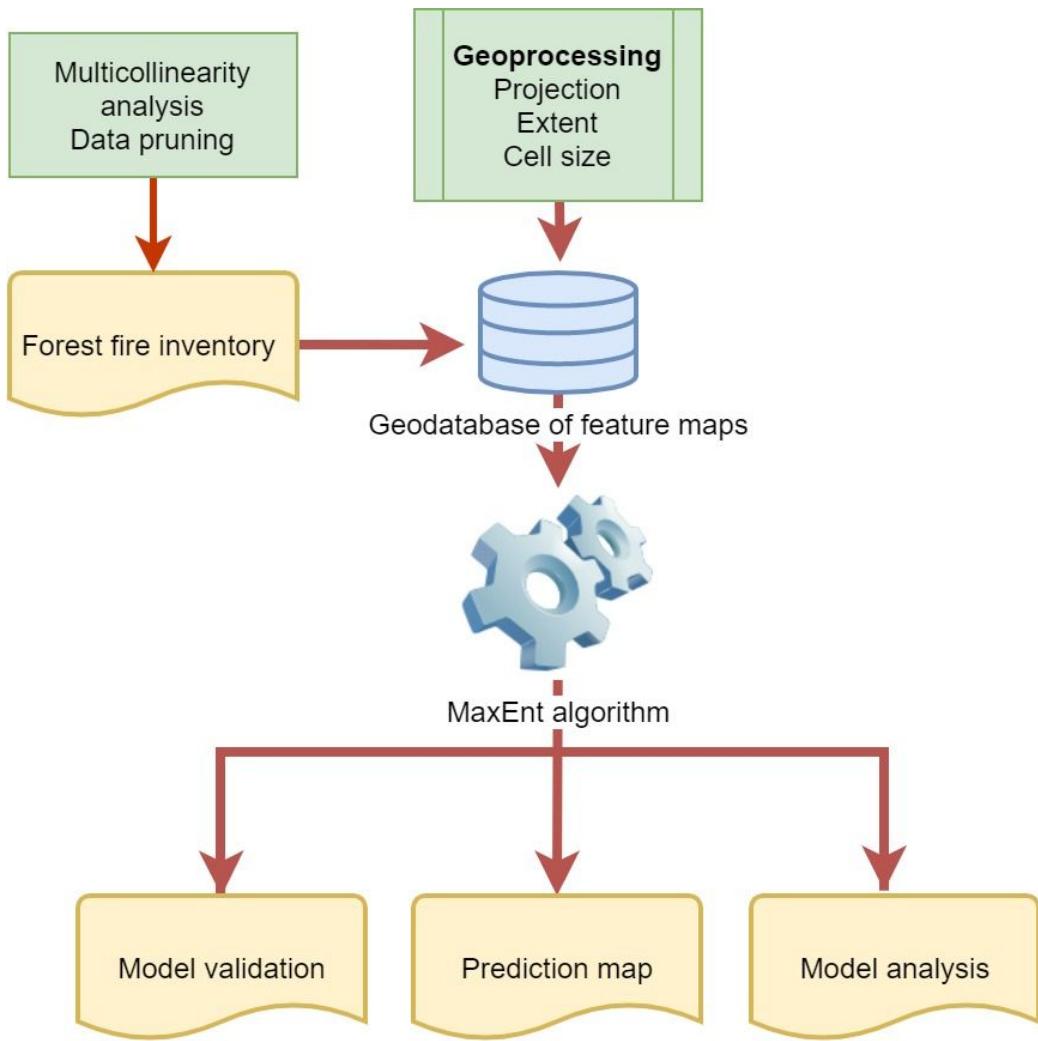


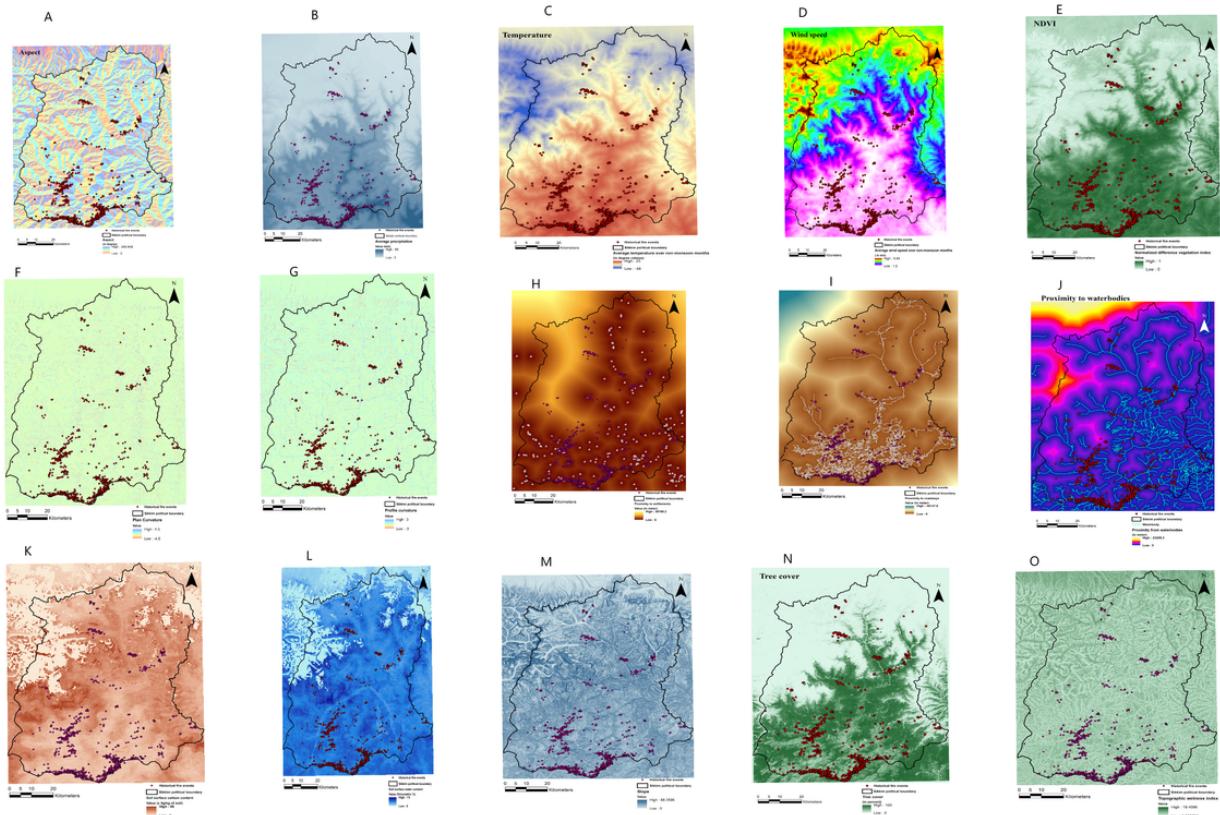
Figure 2

Forest fire trend analysis. Holt's forecast model was used to forecast the likelihood of forest fire events to the year 2022. The forecast has an average boundary of  $\pm 62.343$  wildfire events from 2020 onwards.



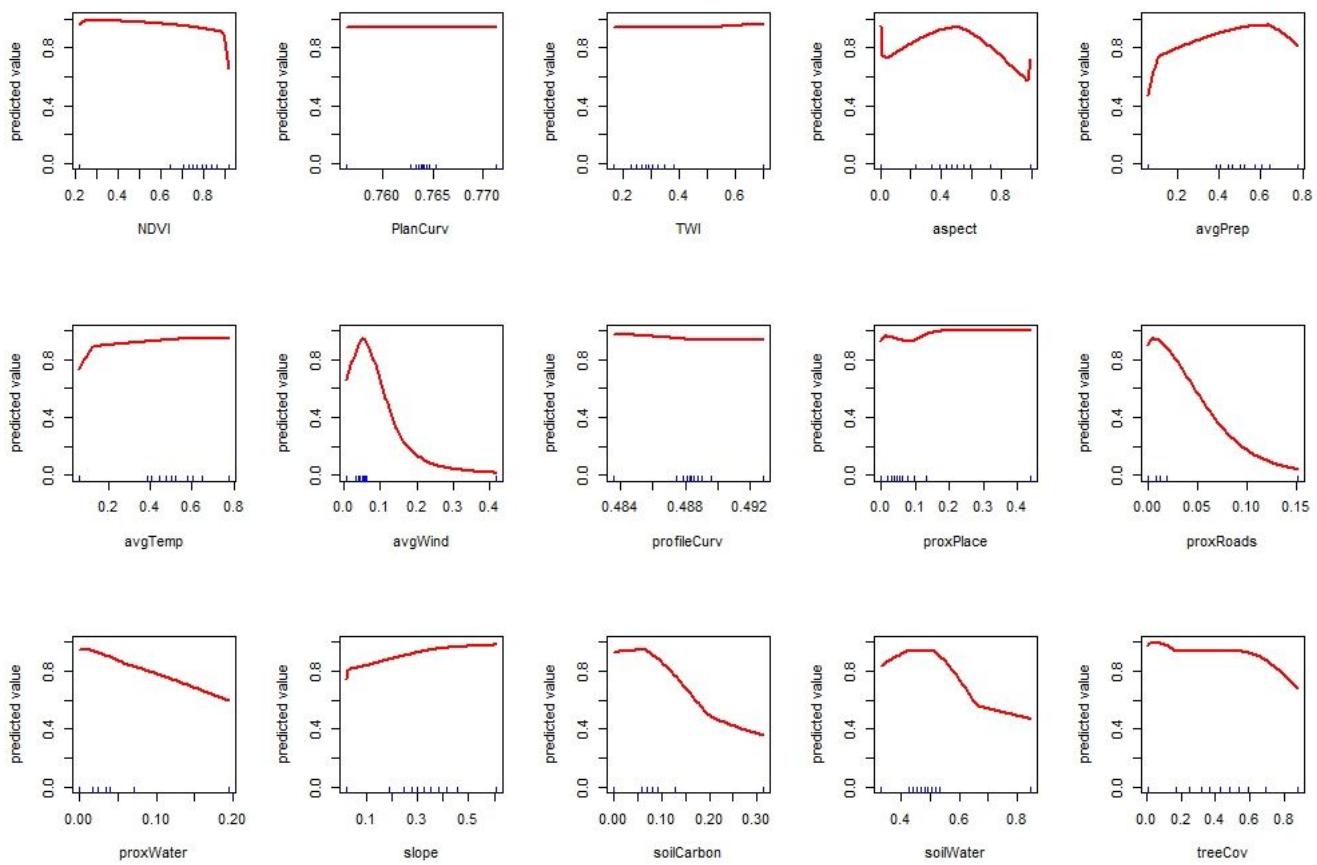
**Figure 3**

Methodology of the model. Initially geoprocessing of the raster feature dataset and forest fire inventory was performed. The processed dataset was stored in a geodatabase and exported as GeoTiff files for analysis in the R environment. MaxEnt based prediction mapping was performed followed by model validation and analysis.



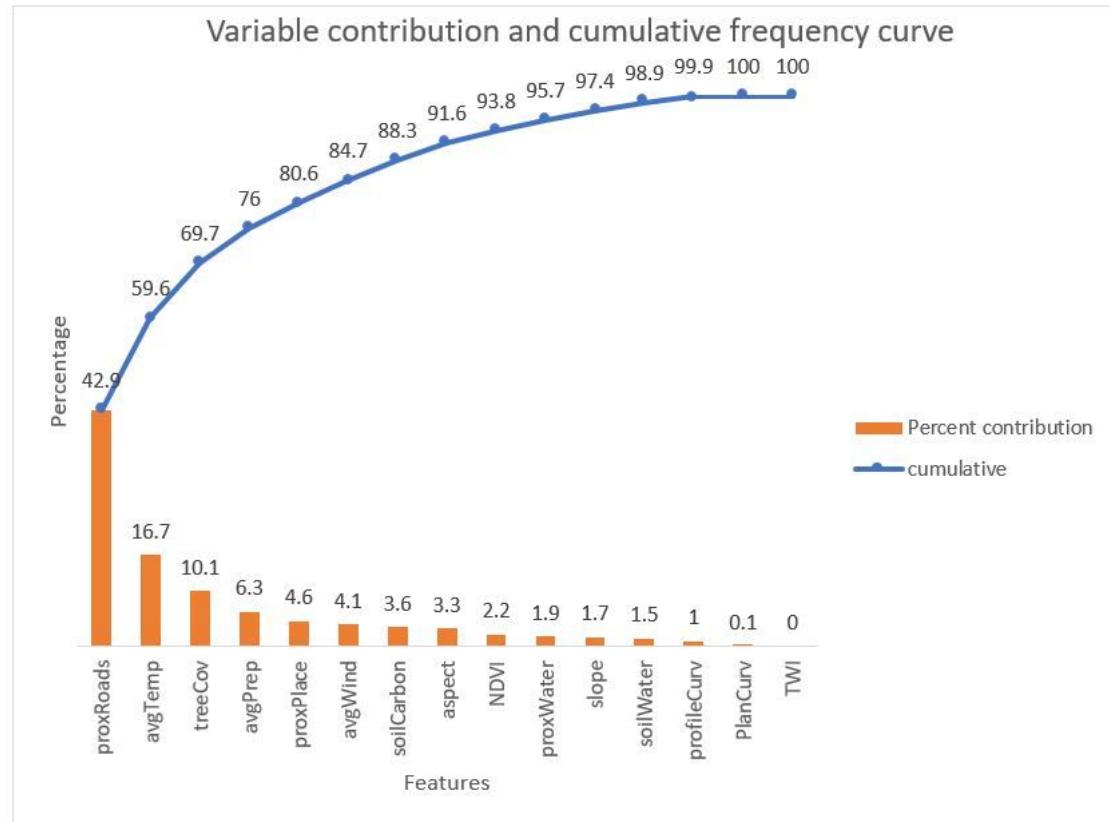
**Figure 4**

Environmental features for the training of MaxEnt. (a) Average ambient temperature in °C . (b) Average rainfall in mm/year. (c) Average wind speed in m/s. (d) Elevation in m. (e) Slope in degree. (f) TWI as a nondimensional value. (g) Aspect in degree. (h) NDVI a nondimensional value. (i) Tree cover in %. (j) Proximity to waterbodies in m. (k) Proximity to human habitations in m. (l) Proximity to roadways in m. (m) Population density in household/km<sup>2</sup>. [No caption provided for figure o.]



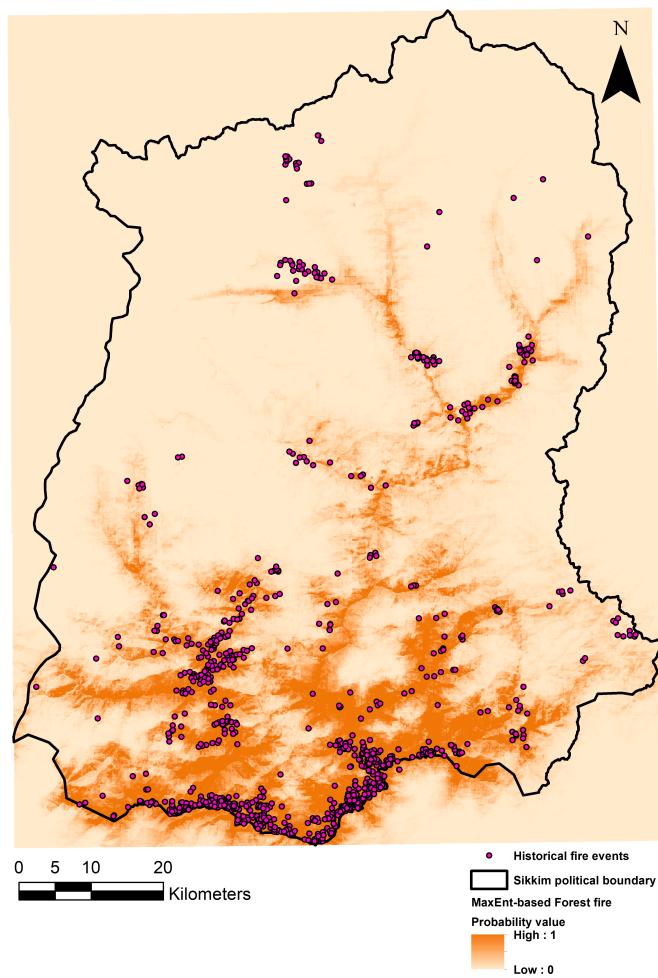
**Figure 5**

Response of the MaxEnt prediction to the feature values.



**Figure 6**

Importance of feature variables.



**Figure 7**

Spatial probability distribution map of forest fire over the study area based on the estimations of MaxEnt model.

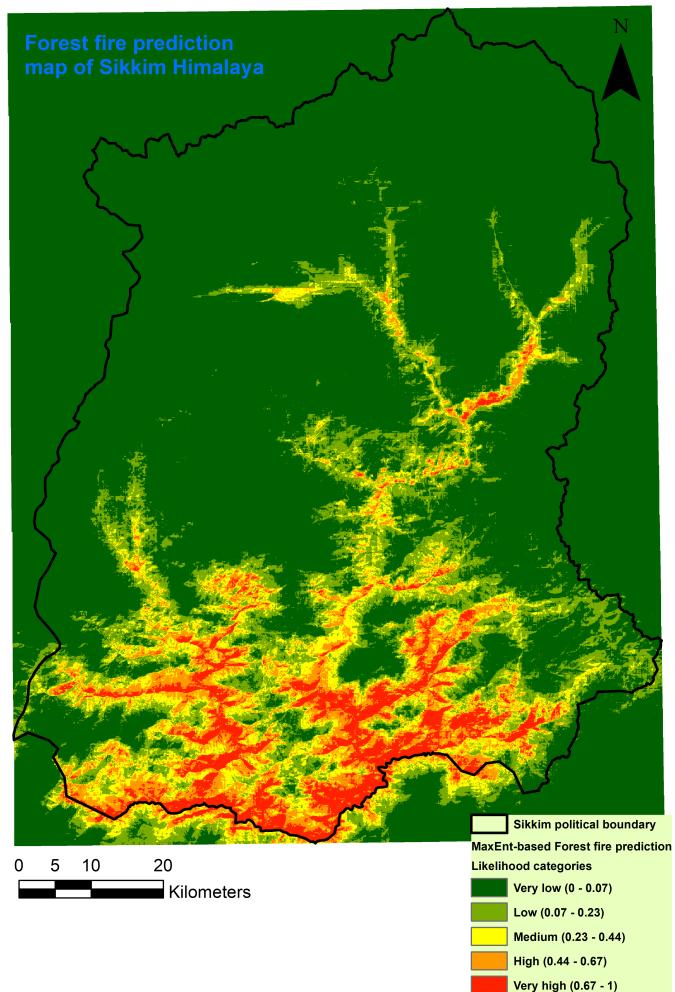
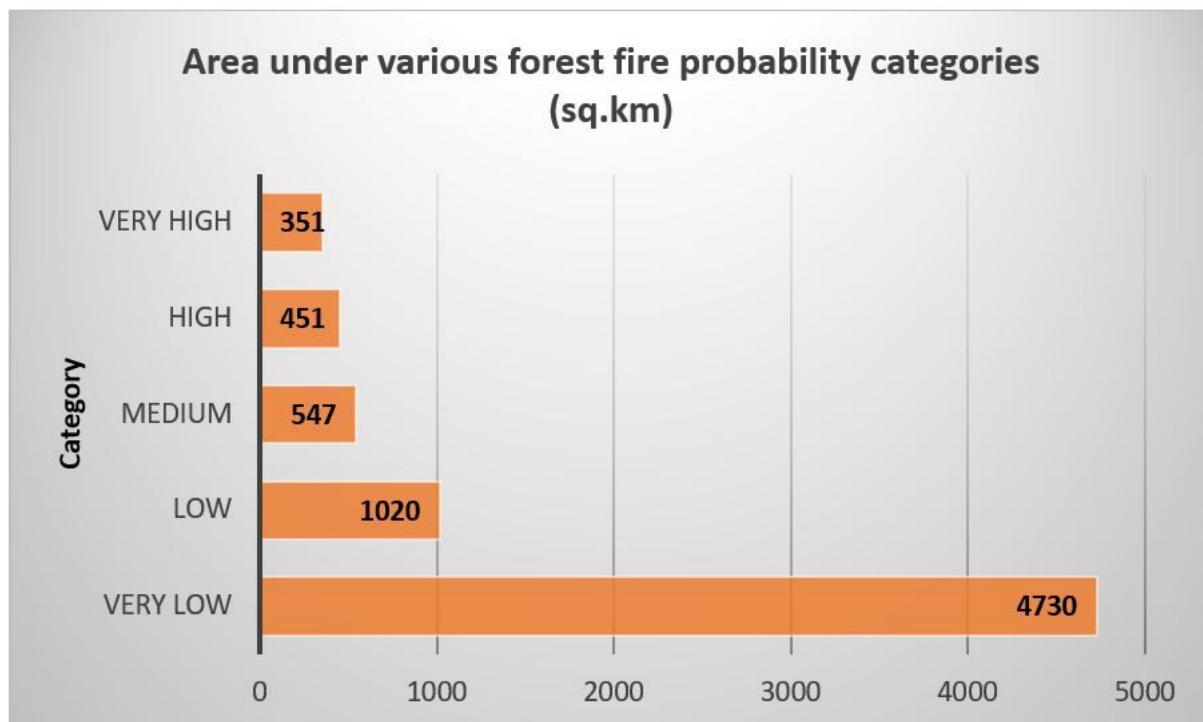


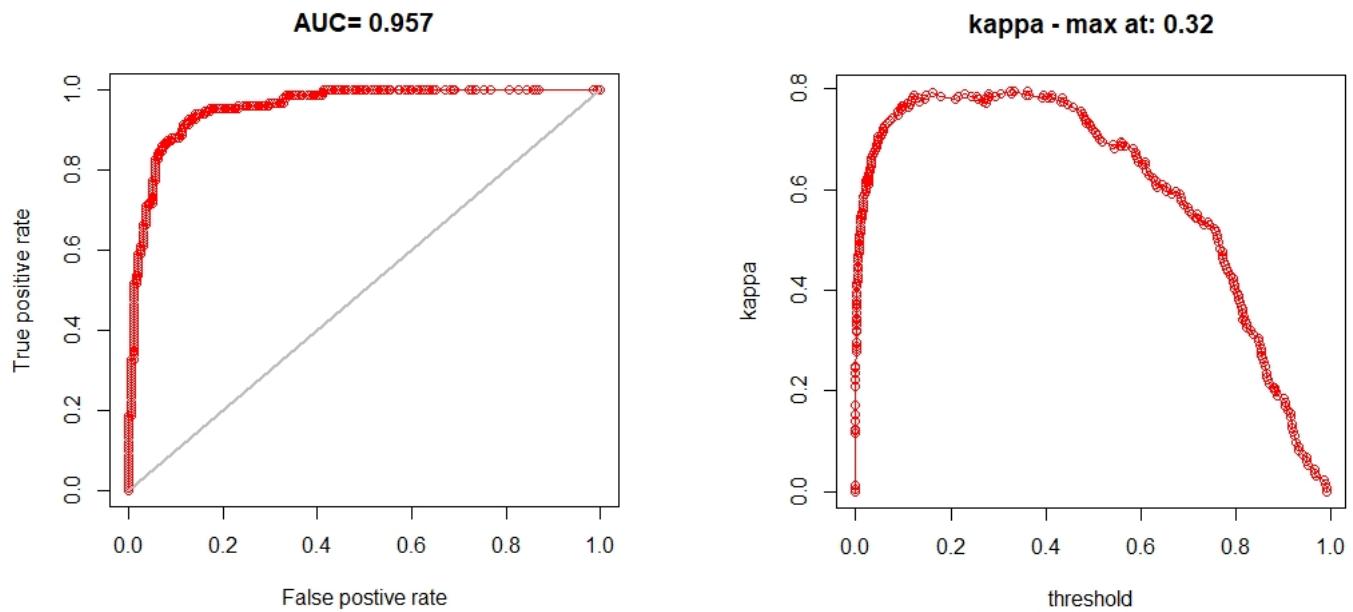
Figure 8

MaxEnt-based categorised forest fire prediction map of Sikkim Himalaya.



**Figure 9**

Forest fire likelihood category wise distribution of the study area.



**Figure 10**

Model validation. (a) ROC curve. (b) Cohen's Kappa curve.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplement.docx](#)