Forest Fire Susceptibility Zonation using dNBR and Machine Learning models: A case study at the Similipal Biosphere Reserve, Odisha, India

Rajkumar Guria
rkguria.007@gmail.com

Fakir Mohan University
https://orcid.org/0000-0002-4499-0283

Manoranjan Mishra
Fakir Mohan University

Samiksha Mohanta
Fakir Mohan University

Suman Paul
Fakir Mohan University

Research Article

Keywords: Forest Fire Susceptibility, Machine learning, Biodiversity, dNBR, Similipal Biosphere Reserve

Posted Date: July 16th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4344777/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Forests play a pivotal role in maintaining environmental equilibrium, chiefly due to their biodiversity. This biodiversity is instrumental in atmospheric purification and oxygen production. Nowadays forest fires are an exciting phenomenon, identification of forest fire susceptible (FFS) areas is necessary for forest fire mitigation and management. This study delves into forest fire trends and susceptibility in the Similipal Biosphere Reserve (SBR) over the period of 2012–2023. Utilizing four machine learning models such as Extreme Gradient Boosting Tree (XGBTree), AdaBag, Random Forest (RF), and Gradient Boosting Machine (GBM). Forest fire inventory was prepared using the Delta Normalized Burn Ratio (dNBR) index. Incorporating 19 conditioning factors and rigorous testing for collinearity, FFS maps were generated, and finally, model performance was evaluated using ROC-AUC, MAE, MSE, and RMSE methods. From the results, it was observed that, overall, about 33.62% of the study area exhibited high to very high susceptibility to forest fires. RF exhibiting the highest accuracy (AUC = 0.85). Analysis of temporal patterns highlighted a peak in fire incidents in 2021, particularly notable in the Buffer Zone. Furthermore, a significant majority (94.72%) of fire incidents occurred during March and April. These findings serve as valuable insights for policymakers and organizations involved in forest fire management, underscoring the importance of targeted strategies for high-risk areas.

1. Introduction

Forests are essential guardians of our planet's health, preserving biodiversity and regulating the climate by serving as a carbon sink. Their diverse plant life also contributes significantly to air purification (Moayedi et al., 2020; Abid, 2021; Bera et al., 2022). The event of forest fires escalating worldwide concerning several natural and anthropogenic phenomena such as climate change, global warming, urbanization, timber trade, and agricultural expansion (Saha et al., 2020). Furthermore, anthropogenic activity plays a significant contributor to forest degradation in forest fires, more than natural activity (Sachdeva et al., 2018; Rajan and Shanmugam, 2018). It was accounted that in recent decades, forest fire events have occurred more frequently with greater severity, impacting 420 million hectares of land each year (Tien Bui et al., 2019; Giglio et al., 2018). Around 90% of forest fire phenomena happen due to anthropogenic while 10% of forest fire phenomena occur naturally (Jain et al., 2020). The severe consequences of forest fires, adversely impact the health of soil and water, resulting in land degradation and erosion, which in turn disturbs the balance of the ecosystem (Mohajane et al., 2021).

India has faced different challenges in forest fire management notably the higher incidence of forest fire phenomena particularly compared to other Southeast Asian nations (Bar et al., 2020). In India, anthropogenic activity is responsible for over 95% of forest fire activities (Jain et al., 2021). The state of Odisha faces significant vulnerability and nearly half of the forest area in Odisha is classified as highly or moderately fire-prone (FSI, 2021). The 2020–2021 period saw a peak in fire incidents in Odisha, particularly in the Angul and Kandhamal districts (FSI, 2021). The continual wildfires persistent a continual threat to forested areas of (Das et al., 2023). The Simlipal Biosphere Reserve (SBR) in Odisha, stands out as a significant hotspot for forest fires, facing annual forest incidents (Ranjan et al., 2023).

The forest fire incident escalates adverse effects on biodiversity and socioeconomic aspects and needs nuanced prevention and suppression strategies (Nami et al., 2018). Nowadays there has been an emphasis on advancements focused on developing comprehensive databases and maps for pinpointing forest fire risk areas. Contemporary the endeavours harness power of utilizing high-resolution satellite data and sophisticated machine learning (ML) methods to enhance the precision of Forest Fire Probability (FFP) mapping (Tien Bui et al., 2019; Moayedi et al., 2020; Mohajane et al., 2021; Bera et al., 2022; Dhar et al., 2023; Das et al., 2023). Satellite data, utilized as a
dependent variable in various research endeavours, play a crucial role in facilitating forest fire susceptibility maps (Gigović et al., 2019; Mohajane et al., 2021; Abid, 2021; Bera et al., 2022; Dhar et al., 2023). While the MODIS and VIIRS data provide high temporal frequency, their coarse spatial resolution restricts accurate monitoring at the micro-scale level (Bar et al., 2020). Medium to high-resolution data such as Sentinel-2 MSI data, which provides enhanced capabilities in the near-infrared (NIR) and short-wave infrared (SWIR) bands for detecting burned and unburned areas (Shi and Zhang, 2023; Llorens et al., 2021; Konkathi and Shetty, 2021). Additionally, the availability of 30m spatial resolution data from the Landsat satellite since June 20, 2022 (Giglio et al., 2018), further enriches the resources for accurate forest fire monitoring.

A diverse array of ML models, such as Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes Tree (NBT), Generalized Additive Model (GAM), Multivariate Adaptive Regression Splines (MARS), and Artificial Neural Network (ANN), have been utilized for creation and predicting the FFP maps (Dimuccio et al., 2011; Pourtaghi et al., 2016; Jaafari et al., 2018; Jaafari and Pourghasemi, 2019; Tien Bui et al., 2019; Bera et al., 2022). The perpetual improvement of forest fire assessment techniques is vital crafting in the formulation of prevention strategies. In Odisha, while Bera et al. (2022) identified fire-prone areas in the SBR region, their study was constrained by the limited resolution of the MODIS dataset, underscoring the need for more precise data (Vabalas et al., 2019). Das et al. (2023) advanced this research by employing various MCDM methodologies for predicting forest fire-prone zones.

The study employs the delta Normalized Burn Ratio (dNBR) techniques using medium to high-resolution Sentinel-2 data to accurately map burn areas. The study integrates Sentinel-2, MODIS, SRTM-DEM, GLOBAL-PET and ESRI-LULC data for identifying the conditional factors of a forest fire. The nobility of the study, focused on trend and pattern analysis of forest fires annually, monthly and spatially and also assesses the forest fires' susceptibility zones by employing the advancement ML methods such as XGBTree, RF, Adabag and GBM, integrating the mitigation and management plant.

2. Material and methods

2.1 Study area

Similipal Biosphere Reserve (SBR) is a significant ecological area within the Northern Eastern Ghats situated between 21°15′ N and 22°25′ N, and 86°00′ E and 86°45′ E (Fig. 1). Covering an expansive area of 5,569 km². It comprises core, buffer, and transition zones, hosting diverse ecosystems and wildlife, including apex predators like tigers. The region's undulating terrain, with elevations ranging from 200 to 1168 meters above sea level, supports various vegetation types and contributes to its ecological richness. The warm and humid climate, with average annual rainfall between 1200 mm to 2000 mm and temperatures ranging from 9°C to 33.5°C, sustains the reserve's biodiversity. SBR is home to numerous tribal communities, accounting for about 75% of the total population, who rely on forest resources for their livelihoods. Recognized as one of India's Mega Biodiversity Areas, SBR was included in UNESCO's Man and Biosphere Program in 2009 due to its significant ecological importance.

2.2 Databases:

In this study, various datasets including Sentinel-2 Multi-Spectral Instrument (MSI) images, SRTM DEM, active fire point data, and field data. Table 1 provides a detailed description of the data sources.
Table 1  
Data sources for forest fire susceptibility zonation mapping

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Prepared from</th>
<th>Data format</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active fire point data</td>
<td>FIRMS</td>
<td>Vector</td>
<td><a href="https://firms.modaps.eosdis.nasa.gov">https://firms.modaps.eosdis.nasa.gov</a></td>
</tr>
<tr>
<td>Slope, TWI, TRI</td>
<td>SRTM DEM</td>
<td>Raster</td>
<td>(<a href="https://earthexplorer.usgs.gov">https://earthexplorer.usgs.gov</a>)</td>
</tr>
<tr>
<td>Rainfall, Relative humidity, Temperature, Solar radiation, Wind speed</td>
<td>NASA power data</td>
<td>Vector</td>
<td>(<a href="https://power.larc.nasa.gov/">https://power.larc.nasa.gov/</a>)</td>
</tr>
<tr>
<td>Potential evapotranspiration</td>
<td>Global-PET</td>
<td>Raster</td>
<td>(<a href="https://figshare.com/">https://figshare.com/</a>)</td>
</tr>
<tr>
<td>Land surface temperature (LST)</td>
<td>MODIS</td>
<td>Raster</td>
<td>MODIS ‘MOD11A1.061’ and ‘MOD17A2H.061’ products</td>
</tr>
<tr>
<td>MNDWI, NDMI, NDVI</td>
<td>Sentinel–2MSI</td>
<td>Raster</td>
<td>Open Access Hub (<a href="https://scihub.copernicus.eu/">https://scihub.copernicus.eu/</a>)</td>
</tr>
<tr>
<td>Distance from cropland, Distance from settlement</td>
<td>Extracted from LULC</td>
<td>Vector</td>
<td>ESRI LULC (<a href="https://www.esri.com">https://www.esri.com</a>)</td>
</tr>
<tr>
<td>Distance from road</td>
<td>Road layer</td>
<td>Vector</td>
<td><a href="https://www.openstreetmap.org/">https://www.openstreetmap.org/</a></td>
</tr>
<tr>
<td>Distance from river</td>
<td>Prepared from SRTM DEM</td>
<td>Vector</td>
<td>ESRI LULC (<a href="https://www.esri.com">https://www.esri.com</a>)</td>
</tr>
<tr>
<td>Distance from surface water bodies</td>
<td>Extracted from MNDWI</td>
<td>Vector</td>
<td>ESRI LULC (<a href="https://www.esri.com">https://www.esri.com</a>)</td>
</tr>
<tr>
<td>Land use and land cover (LULC)</td>
<td>ESRI LULC</td>
<td>Raster</td>
<td>(<a href="https://www.esri.com">https://www.esri.com</a>)</td>
</tr>
<tr>
<td>NBR, dNBR</td>
<td>Sentinel–2MSI</td>
<td>Raster</td>
<td>Open Access Hub (<a href="https://scihub.copernicus.eu/">https://scihub.copernicus.eu/</a>)</td>
</tr>
</tbody>
</table>

2.3 Demarcation of burn area by delta Normalized Burn Ratio (dNBR)

The delta normalized burn ratio (dNBR) is a valuable tool for understanding forest fires. It uses data from the near-infrared (NIR) and short-wave infrared (SWIR2) spectral regions to differentiate between burned and unburned areas (Llorens et al., 2021). By focusing on areas less impacted by atmospheric effects, dNBR accurately measures the effects of fire on vegetation and canopy moisture loss (Soverel et al., 2010). This makes it useful for assessing post-fire environmental changes, and capturing features of burned landscapes while minimizing interference from atmospheric conditions. The dNBR relies on pre-fire and post-fire data to calculate changes in vegetation health (Miller and Thode, 2007; Shi and Zhang, 2023). The steps outlined to determine the demarcation of burn areas are as follows:

In our study, we analyzed imagery captured by the Sentinel-2 satellite’s multispectral instrument (MSI). We used Green band–3 (560 nm) at 10 m resolution, Near-Infrared (NIR) band–8 (842 nm), and Short-Wave Infrared (SWIR2) band–12 (2190 nm) at 20 m resolution. To ensure consistency, we resampled the SWIR2 imagery from 20 to 10 m resolution using bicubic resampling in Google Earth Engine (Liu et al., 2020).

2.3.1 Normalized Burn Ratio (NBR)
The NBR from Sentinel-2 MSI imagery to identify burned and unburned areas (Miller and Thode, 2007). NBR is calculated using the NIR and SWIR2 bands. NIR detects living plant chlorophyll content, while SWIR2 distinguishes non-photosynthetic dead trees and other materials in post-fire images (Bar et al., 2020). A decrease in NIR reflectance and an increase in SWIR2 reflectance indicate burned areas. NBR calculation followed Eq. (1)

\[
NBR = \frac{\rho_{\text{nir}} - \rho_{\text{swir2}}}{\rho_{\text{nir}} + \rho_{\text{swir2}}}
\]

where \(\text{nir}\) and \(\text{swir2}\) represent band 8 and band 12 of Sentinel–2 respectively. The NBR values ranged between −1 and +1.

By analyzing the active fire point data of the past 12 years (2012 to 2023), we found that the year 2021 had the highest number of forest fire incidents in the region, so we took this year as a sample for determining fire severity. Furthermore, forest fires are most prevalent during the pre-monsoon period, notably from March to April. Therefore, we designated February as the pre-fire period for analysis, while May was chosen as the post-fire period for computing the NBR. This temporal distinction is essential for accurately evaluating the impact of forest fires in the region.

2.3.2 Delta Normalized Burn Ratio (dNBR)

The delta normalized burn ratio (dNBR) method is crucial for identifying burn severity areas, offering a reliable approach by analyzing pre- and post-fire changes (Miller and Thode, 2007). This method effectively captures spatial variations in fire intensity within specific boundaries, making it ideal for this study. The dNBR is calculated using the Eq. (2):

\[
dNBR = NBR_{\text{pre fire 2021}} - NBR_{\text{post fire 2021}}
\]

Negative dNBR values indicate vegetation regeneration, while positive values signify burned areas. The United States Geological Survey (USGS) defines thresholds for dNBR ranging from −0.5 to +1.3, with seven distinct classes provided for interpretation (Suresh Babu et al., 2018).

2.3.3 Mitigating Water Pixel Interference

Water bodies can complicate assessing fire severity before and after fires because water and burnt areas can look similar in satellite images. To fix this, researchers used the Normalized Difference Water Index (NDWI) to accurately find and remove water pixels (Llorens et al., 2021). This helped make the burnt area maps more accurate and reduced errors. Removing water improved the precision of assessing fire severity and reduced confusion between water and burnt areas in the analysis (Suresh Babu et al., 2018; Llorens et al., 2021).

2.4 Generation of Forest Fire Inventory

Various studies have used MODIS and VIIRS satellite data to create forest fire susceptibility maps (Gigović et al., 2019; Mohajane et al., 2021; Abid, 2021; Bera et al., 2022; Dhar et al., 2023), but these datasets might not provide precise results for monitoring fires on micro-scale regions (Bar et al., 2020). For more accurate and detailed identification of fire areas, medium to high-resolution data is needed. Sentinel-2, a Multi-Spectral Instrument (MSI) satellite, offers both medium and high spatial resolution (10 – 20 m) in the NIR and SWIR bands (Shi and Zhang,
This capability is crucial for effectively detecting both burned and unburned areas (Llorens et al., 2021). A forest fire inventory map is crucial for understanding past fire patterns and identifying areas prone to future fires. In our study, we created a fire map using dNBR calculations. We divided dNBR values into two sets based on a threshold (≥ 0.1): one set representing burned pixels (1000 fire points) and the other representing unburned pixels (1000 non-fire points). By using ArcGIS v10.4’s “Create Random Points” tool, we collected a total of 2000 points. To facilitate model training and validation, we employed a random-partition algorithm, allocating 70% of the data for training and 30% for testing (Piao et al., 2022).

### 2.5 Selection and Collinearity assessment of FF conditioning factors

We analyzed the selected factors contributing to fire initiation using the Variance Inflation Factor (VIF) to check for multicollinearity among the variables. Tolerance was also used for this purpose (Guria et al., 2024). Additionally, we used Pearson correlation to assess how closely related the predictor variables were spatially. The 19 influencing factors for forest fires examined in the current study have been categorized into four main groups: physical, climate, biophysical, and anthropogenic factors (Fig. 2). The Topographic Roughness Index (TRI), Topographic Wetness Index (TWI), and slope layers were generated using the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) with a spatial resolution of 30 meters (Fig. 3). Slope position influences fire behavior, impacting flame lengths and rates of spread (Pourtaghi et al., 2016). TRI and TWI reflect topographic influence and are tied to soil conditions. As a measure of topographic complexity, it indirectly relates to forest fires (Fang et al., 2018). Hot, high wind speeds, and dry weather, influenced by factors such as temperature, solar radiation, wind speed, precipitation, PET, and relative humidity, strongly contribute to the behavior of forest fires, increasing the likelihood and intensity of fires (Mhawej et al., 2015). The Normalized Difference Moisture Index (NDMI) and Modified Normalized Difference Water Index (MNDWI) serve as spectral indices for mapping forest fire probability. NDMI evaluates vegetation water stress or content, aiding in the prediction of fire risk (Bera et al., 2022). Conversely, MNDWI assesses water content in vegetation, potentially influencing forest fire susceptibility. The Normalized Difference Vegetation Index (NDVI) provides valuable insights into vegetation health, playing a pivotal role in assessing the risk of forest fires (Meng et al., 2015). Furthermore, increased Land Surface Temperature (LST) can lead to drier environments, reducing soil moisture content and consequently raising the susceptibility of vegetation to catch fire. The influence of human activities on forest fires can be assessed by examining factors like the distance from roads, settlements, and land use patterns (Kalantar et al., 2020). Roads, serving as pathways for human interaction with forested regions, may escalate the risk of accidental or deliberate fire ignition (Sari, 2021; Mohajane et al., 2021). Similarly, the proximity of settlements, agricultural areas, and various land uses, including residential and industrial zones, can heighten the vulnerability to fires. Investigating these aspects provides insights into how human actions and infrastructure affect the occurrence, propagation, and control of forest fires (Tien Bui et al., 2019; Chamling and Bera, 2020). The proximity to surface water bodies can significantly influence forest fire susceptibility by affecting moisture availability in vegetation. Areas closer to rivers, lakes, or other water bodies tend to have higher humidity levels, which can reduce the likelihood and intensity of forest fires (Pourtaghi et al., 2016). Areas in close proximity to water sources tend to have higher moisture availability, particularly those near rivers, resulting in elevated humidity levels. This increased humidity can reduce the likelihood and intensity of forest fires (Ngoc Thach et al., 2018).

### 2.6 Methods for Forest Fire Probability Mapping

#### 2.6.1 AdaBag
In our forest fire probability prediction model, we carefully selected the fitting parameters for AdaBag to optimize its performance. First, we chose $m_{final} = 10$, allowing for the aggregation of up to 10 weak learners to create a robust combined model. By setting $coefficient = \text{Freund}$, we utilize the Freund and Schapire, (1997) learning coefficient, known for its effectiveness in boosting weak learners (Wang et al., 2019). Given the critical nature of forest fire prediction, we disabled boosting techniques by setting $boos = \text{FALSE}$ to focus on other ensemble methods. We constrained individual decision trees’ complexity by setting $maxdepth = 5$, ensuring they don’t become overly intricate and prone to overfitting. Setting $minsplit = 0$ allows nodes to split even with minimal observations, ensuring the algorithm captures potentially crucial patterns in the data. For the complexity parameter ($cp$), we opted for the default value (-1), allowing AdaBag to automatically determine the optimal tree complexity. Lastly, enabling sampling with replacement ($replace = \text{TRUE}$) enhances the diversity of weak learners, improving the model’s robustness and predictive accuracy in forest fire probability mapping.

2.6.2 XGBTree

To predict forest fire probability using XGBoost, we meticulously selected the fitting parameters to optimize its performance. First, by setting $max_depth = 2$, we limited the maximum depth of each tree in the ensemble, preventing overly complex models and mitigating overfitting. The parameter $eta = 1$ specifies the learning rate, controlling the step size during the optimization process. With $nthread = 2$, we utilized two CPU threads for parallel processing, enhancing computational efficiency. To ensure the robustness of our model evaluation, we employed 5-fold cross-validation with $nfold = 5$, splitting the data into five subsets for training and testing. By specifying $nrounds = 200$, we set the number of boosting rounds to 200, determining the number of trees to be built. Additionally, $verbose = \text{TRUE}$ enables verbose output, providing detailed information about the training process. The metric = ROC indicates that we used the Receiver Operating Characteristic (ROC) curve as the evaluation metric, assessing the model’s ability to distinguish between positive and negative instances. We performed 10 iterations of the algorithm using $number = 10$, refining the model iteratively. Finally, we applied regularization techniques by setting $lambda = 0.0005$ and $alpha = 0.0005$, controlling Ridge and Lasso regularization, respectively, to prevent overfitting and improve generalization performance in predicting forest fire probability.

2.6.3 Gradient Boosting Machine (GBM)

In GBM we initiated the process with ‘method = gbm’ to specify the use of the GBM algorithm. By setting ‘number = 10’, we performed 10 iterations of the algorithm, repeating the process 10 times (‘repeats = 10’) for robustness. To streamline the output, we chose ‘verbose = FALSE’. ‘classProbs = TRUE’ enabled the model to calculate class probabilities, providing additional insights. We utilized ‘summaryFunction = two’ to aggregate results using a two-category summary function. ‘tuneGrid = gbmGrid’ facilitated hyperparameter tuning using predefined grid parameters. Finally, ‘metric = ROC’ indicated our evaluation metric as the Receiver Operating Characteristic (ROC) curve, assessing the model’s ability to distinguish between positive and negative instances accurately.

2.6.4 Random Forest (RF)

To facilitate our analysis, we employed the ‘randomForest’ and ‘caret’ libraries in R. Setting the random seed to 1234 ensured the reproducibility of our results across multiple runs. We meticulously crafted our model’s training process by implementing a 10-fold cross-validation strategy using the ‘cv’ method. Furthermore, we fine-tuned our RF model by exploring various configurations of the ‘mtry’ parameter, which governs the number of variables considered at each split in the RF algorithm. Through an exhaustive grid search ranging from 2 to 15 for ‘mtry’, we systematically identified the optimal hyperparameters that maximized the model’s predictive accuracy. Additionally, we incorporated a preprocessing step to standardize feature scaling across the dataset, ensuring uniformity in the input
variables. Finally, our RF model was trained on the training dataset, with the selected hyperparameters and 100 to 500 for 'ntree'.

2.7 Validation of dNBR and ML models

In our accuracy assessment, we compared our burn severity map (dNBR) with VIIRS active fire point data. Burned areas with severity equal to or greater than 0.1 were labeled as 1, while unburned areas were labeled as 0. By overlaying fire points onto the map, we verified if they matched the corresponding burn severity class. We used four statistical methods (MSE, MAE, RMSE, and Overall accuracy) to measure agreement between observed and predicted fire points. These assessments, informed by equations (3) to (5), helped evaluate the reliability and precision of our mapping approach (Rasool et al., 2022; Pan et al., 2022).

\[
MSE = \frac{\sum_{i=1}^{n} (X_i - Y_i)^2}{n}
\]

3

\[
MAE = \frac{\sum_{i=1}^{n} |X_i - Y_i|}{n}
\]

4

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_i - Y_i)^2}{n}}
\]

5

where, \( n \) depicted number of fire points, \( X_i \) and \( Y_i \) represents number of observed and predicted fire pixel respectively, \( Y \) is total number of fire pixel for the \( k \)-th class.

The Receiver Operating Characteristic (ROC) curve, a common method for assessing predictive models (Guria et al., 2024), with data from dNBR maps to evaluate four machine learning models. Additionally, we employed MAE, MSE, and RMSE for further model validation, the AUC was calculated using the Eq. (6).

\[
AUC = \sum_{k=1}^{n} \left( M_{K+1} - M_K \right) \left( N_K + 1 - N_{K+1} - \frac{N_K}{2} \right)
\]

6

3. Results and analysis

3.1 Analyzing trends and patterns of forest fire in SBR

In the current study, the temporal trend and pattern analysis of forest fires were conducted employing VIIRS active fire point data for the periods of 2012–2023. From Fig. 4a, it is observed that there has been an upward trend of forest fires in the SBR region during the last 12 years. The highest number of fire incidents was observed in 2021 (\( n = 4349 \)), with 2018 (\( n = 2596 \)) and 2013 (\( n = 2224 \)) also experiencing high numbers of fires. Figure 4b shows the graphical representation of the monthly forest fire count for each year.
Figure 4c displays forest fire incidents across different zones from 2012 to 2023. In the Buffer Zone, fire incidents fluctuated over the years, reaching a peak of 2417 fires in 2021 and a low of 45 fires in 2019, totaling 9285 fires over the period. Similarly, the Core Zone experienced variable fire counts, with the highest recorded in 2021 (1150 fires) and the lowest in 2019 (48 fires), summing up to 5848 fires. In the Transitional Zone, fire occurrences varied, peaking at 782 fires in 2020 and hitting a low of 47 fires in 2019, with a total of 2740 fires observed throughout the 12 years.

During the study period, the analysis reveals that approximately 94.72% of the fire incidents occurred in the months of March and April combined. Specifically, March accounted for a significant portion, with 73.42% of the total fire incidents recorded during this period. As a result, March and April are identified as the primary fire season based on the concentration of fire incidents (Fig. 4d). Forest fires mainly occur in this region during the pre-monsoon season.

3.2 Analysis of dNBR and Accuracy Assessment

In the present study, burn severity assessment was performed using the NBR and the dNBR indices for the year 2021 (Fig. 5). These indices were utilized to categorize burn severity into five distinct levels: Unburned (dNBR values below 0.1), Low Severity (dNBR values ranging from 0.1 to 0.269), Moderate-Low Severity (0.270 to 0.439), Moderate-High Severity (0.440 to 0.659), and High Severity (0.660 to 1.3). The analysis of burn severity using the dNBR indices for the year 2021 revealed distinct categories of fire impact across the study area. The majority of the area, accounting for 73.42%, fell within the Low Severity and High Severity categories, indicating varying degrees of fire damage. Specifically, Low Severity areas covered 51.08% of the total area, while High Severity areas comprised 22.53%. Moderate-Low and Moderate-High Severity categories contributed to 7.67% and 6.66% of the area, respectively. Unburned areas constituted a smaller proportion at 12.06%.

The predictive capability assessment for the dNBR index using various accuracy measurements for the year 2021. MSE is calculated as 0.142, indicating the average squared differences between the predicted and observed values. The MAE is reported as 0.183, representing the average absolute differences between the predicted and observed values. The RMSE, computed as 0.194, signifies the square root of the average squared differences between the predicted and observed values. Finally, the Overall Accuracy is determined to be 0.883, indicating the proportion of correctly classified samples out of the total samples assessed. These accuracy metrics provide valuable insights into the performance of the dNBR index in predicting burn severity levels.

3.3 Multi-collinearity
Table 2 presents the collinearity statistics for various parameters used in the analysis. Collinearity is assessed using two metrics: Tolerance and Variance Inflation Factor (VIF). Tolerance values closer to 1 indicate low collinearity, while VIF values above 10 suggest high collinearity. Among the parameters, Rainfall exhibits the lowest tolerance (0.145), indicating potential collinearity issues, supported by its high VIF value of 6.278. Similarly, Relative Humidity (RH) and Slope also show relatively low tolerances (0.229 and 0.109 respectively) and high VIF values (5.581 and 8.727 respectively). On the other hand, parameters like LULC and MNDWI demonstrate higher tolerances (0.669 and 0.281 respectively) and lower VIF values (1.457 and 3.564 respectively), suggesting lower collinearity with other variables. Overall, the assessment highlights no multicollinearity problem.

### 3.4 Analysis of Forest Fire Susceptibility:

Four machine-learning models, including XGBTree, GBM, AdaBag, and RF, were employed to generate FFP maps, each producing probability predictions (Fig. 6a–d). These probability predictions were categorized into five classes—very high, high, moderate, low, and very low—using the natural-break classification method, a widely accepted approach in classification studies (Xu et al., 2012; Razavi Termeh et al., 2018; Pourghasemi et al., 2019).

According to the XGBTree model, the forest fire susceptibility classes ranged from 17.96% for very low to 16.89% for very high. This implies that approximately 33.75% of the study area had a high to very high susceptibility to forest fires. Similar results were obtained from other models, including AdaBag, RF, and GBM. Across all models, forest fire susceptibility varied, with percentages ranging from approximately 17.00–30.80% for AdaBag, 16.52–24.18% for RF, and 16.94–24.95% for GBM. Overall, about 33.62% of the study area exhibited high to very high susceptibility to forest fires, indicating a significant risk posed by this phenomenon (Table 3).
Table 3 Distribution of forest fire susceptible zones in SBR

<table>
<thead>
<tr>
<th>Models</th>
<th>XGBTree</th>
<th>AdaBag</th>
<th>RF</th>
<th>GBM</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area</td>
<td>%</td>
<td>Area</td>
<td>%</td>
<td>Area</td>
</tr>
<tr>
<td>Very low</td>
<td>973.53</td>
<td>17.96</td>
<td>1669.73</td>
<td>30.80</td>
<td>1212.66</td>
</tr>
<tr>
<td>Low</td>
<td>1125.96</td>
<td>20.77</td>
<td>1123.73</td>
<td>20.73</td>
<td>1073.67</td>
</tr>
<tr>
<td>Moderate</td>
<td>1491.99</td>
<td>27.52</td>
<td>921.60</td>
<td>17.00</td>
<td>1310.84</td>
</tr>
<tr>
<td>High</td>
<td>913.84</td>
<td>16.86</td>
<td>895.49</td>
<td>16.52</td>
<td>895.24</td>
</tr>
<tr>
<td>Very high</td>
<td>915.47</td>
<td>16.89</td>
<td>810.25</td>
<td>14.95</td>
<td>928.38</td>
</tr>
</tbody>
</table>

Within the SBR region, particularly in the core area, forests such as Daldali, Barehipaniin, Kajhari, Khadkei, and Haladia have been identified as prone to forest fires based on our analysis. Additionally, areas like Karkachia, Brahmanagan, Sarada Bisoi Hill, Chirupada, Hatasahi, and Kuanrpal within the buffer zone are highly susceptible to fires. In the transitional zone, forests such as the Badampahar forest range, Godabhanga, and Duara Suni have also been found to be prone to fires.

3.5 Models validation:

Table 4 presents the goodness of fit statistics for different machine learning (ML) models used to predict Forest Fire Susceptibility (FFS). The metrics assessed include Area Under the Curve (AUC), Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The XGBTree model achieved an AUC of 0.83 (Fig. 7), with MSE, MAE, and RMSE values of 0.14, 0.19, and 0.17 respectively. AdaBag and RF models exhibited slightly better performance, with AUC values of 0.84 and 0.85 respectively, and lower MSE, MAE, and RMSE values. The GBM model showed an AUC of 0.82, with slightly higher error metrics compared to the other models. Overall, the RF model demonstrated the best goodness of fit among the models evaluated for predicting FFS.

4. Discussion

Identifying forest fire-prone areas is commonly done using active fire point data obtained from MODIS and VIIRS satellite sensors (Oliva and Schroeder, 2015; Chuvieco et al., 2019). In the SBR region, research integrating machine learning and active fire data has been conducted (Bera et al., 2022; Singha et al., 2024). While these datasets offer valuable insights, they have limitations in detecting fires within forested areas due to factors like closed canopy cover and the presence of clouds or smoke (Coskuner, 2022). To address this, our study utilizes Sentinel-2's higher
spatial resolution images and the DNBAR index to accurately identify burned and unburned areas (Mallinis et al., 2018; Llamas et al., 2019).

We employed four different models – XGBTree, Adabag, RF, and GBM – and selected nineteen factors related to FFS for analysis. Multi-collinearity tests were conducted to ensure the independence of these factors. Additionally, four validation methods were applied to assess the models’ performance, with results indicating satisfactory robustness (AUC above 80%) in predicting fire occurrence.

Our study identifies buffer and transitional zones as the most re-prone areas, consistent with previous findings (Bera et al., 2022). These zones are susceptible to fires due to increased anthropogenic activity and topographic distribution. Climate conditions during the fire season, mainly from March to May, also influence fire development and spread (Singha et al., 2024; Mishra et al., 2024). Fires in the region often originate from dry deciduous forests, ignited by dead leaves during the pre-monsoon season (Saranya et al., 2014). Additionally, natural causes such as lightning contribute to forest fires in the SBR region, characterized by dry deciduous and moist deciduous forests.

Forest fires pose a severe threat to nature and communities, exacerbated by climate change. The Similipal Biosphere Reserve (SBR), a biodiverse region in India, has witnessed an increase in wildfires in recent years (Saranya et al., 2014; Dash and Behera, 2018). To mitigate these fires, modern scientific methods are essential, with geospatial approaches proving effective in monitoring and policymaking (Thapa et al., 2021). Earth Observation satellite sensors, like Sentinel-2 MSI and Landsat-8, aid in detecting and monitoring fires globally (Hu et al., 2021). However, limitations exist, as highlighted by this study, necessitating the integration of cutting-edge technology and region-specific systems such as EFFIS for Europe and Canada's Wildland Fire Information System (San-Miguel-Ayanz et al., 2013; Tymstra et al., 2020). Community-based fire Management initiatives in Africa and holistic approaches in Australia further address wildfire challenges (Moore, 2019; Gonzalez-Mathiesen et al., 2021). In India, the Indian Forest Fire Response and Assessment System (INFRARS) utilizes geospatial data and GIS technology for efficient forest fire management (Matin et al., 2017). This study aids in identifying risk-prone areas and informing forest management strategies, contributing to theoretical insights on forest fire dynamics (Simon et al., 2004; Moore, 2019).

5. Conclusion

The current study leverages advanced machine learning methods to predict Forest Fire Susceptibility zones within the SBR region. Utilizing medium to high-resolution Sentinel-2 MSI data and employing the dNBR method for identifying burned areas, the research aims to offer a more precise understanding of ground realities. Burn area assessment was conducted for the year 2021, cross-validated using VIIRS sensor-derived active fire point data through various statistical methods to ensure accuracy. Fire points within the burn area were randomly selected as dependent variables for the machine learning models, including XGBTree, Adabag, GBM, and RF. Model performance was evaluated using ROC-AUC, MAE, MSE, and RMSE methods, with RF demonstrating superior accuracy (AUC = 0.85).

Furthermore, temporal analysis of forest fire patterns and trends from 2012 to 2023 revealed notable insights. The highest number of fire incidents occurred in 2021, with fluctuations observed in the Buffer Zone over the years. Notably, approximately 94.72% of fire incidents took place during March and April combined.

While the study provides valuable insights, certain limitations exist that warrant further exploration in future research endeavors. Enhancing the validation process by incorporating ground truth validation points alongside medium-
resolution fire point data could bolster result robustness. Additionally, employing more advanced ensemble machine learning or deep learning techniques in fire probability mapping may yield even more accurate outcomes. These avenues present opportunities for future researchers to enhance the precision and reliability of their studies.

**Declarations**

- **Ethical Approval**

Not applicable.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

- **Consent to Participate**

Not applicable.

- **Consent to Publish**

The author is familiar and agree with the content of this paper.

- **Authors Contributions**

Rajkumar Guria: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing –original draft, Writing – review & editing, Visualization. Manoranjan Mishra: Supervision, Conceptualization, Formal analysis, Investigation, Writing – review & editing. Samiksha Mohanta: Data curation, Writing – review & editing. Suman Paul: Formal analysis, Writing – review & editing.

- **Funding**

This study was carried out without any external financial assistance.

- **Competing Interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

**References**


Figures
Figure 1

Location map of the study area (a) India's map highlighting Odisha and SBR, (b) study area with fire points, elevation, rivers, and lakes.
Figure 2

Methodological flow chart
Figure 3

Forest fire conditioning factors
**Figure 4**

Forest fire pattern and trend analysis
Figure 5

NBR of pre and post-fire events and dNBR map
Figure 6

Forest Fire Susceptibility (FFS) maps
Figure 7

Machine learning model validation using ROC-AUC