Muhammad et al. Fire Ecology (2025) 21:72
https://doi.org/10.1186/s42408-025-00408-w

Fire Ecology

ORIGINAL RESEARCH

Open Access

Machine learning-based forest fire vulnerability assessment in subtropical chir pine forests of Pakistan

Sultan Muhammad[1,2], Syed Moazzam Nizami[1*], Qun ou Jiang[3*], Kaleem Mehmood[2,4], Majid Hussain[1], Shoaib Ahmad Anees[5], Fahad Shahzad[6] and Waseem Razzaq Khan[7]

Abstract

Background Subtropical pine forests dominated by *Pinus roxburghii* in northern Pakistan are increasingly vulnerable to forest fires, posing serious ecological and management challenges. This study (i) identifies the primary environmental and anthropogenic drivers of fire risk, (ii) evaluated the performance of four machine learning models Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost) for fire susceptibility prediction, and (iii) develops a spatially explicit vulnerability map for the Malakand region using 2001–2023 fire occurrence data**.**

Methods Spatial modeling was conducted using MODIS fire products (FIRMS and MCD64A1), along with topographic, climatic, vegetation, and human activity variables. Predictor importance and inter-variable relationships were analyzed prior to model training and evaluation.

Results NDVI, skin temperature, and population density emerged as key predictors. Among the models tested, RF and XGBoost outperformed others, with RF achieving 86.2% accuracy and AUC 95.4, and XGBoost showing 87.2% accuracy and AUC 95.2. Multicollinearity analysis confirmed variable independence (Tolerance > 0.1; VIF < 10).

Conclusion The resulting fire vulnerability map delineated distinct spatial patterns, identifying 6.88% of the study area as highly susceptible to fire. These findings provide actionable insights into targeted fire prevention and sustainable forest management in subtropical Himalayan ecosystems.

Keywords Chir pine forest, Swat region, Forest fire vulnerability, Machine learning, Forest fire, Climate, Vegetation indices

Resumen

Antecedentes Los bosques de pinos subtropicales dominados por *Pinus roxburghii* en el norte de Paquistán, son cada vez más vulnerables a los incendios forestales, lo que implica un serio desafío desde lo ecológico y de manejo. Este estudio (i) identifica los conducentes primarios tanto ambientales como antropogénicos que afectan el riesgo de incendios, (ii) evalúa la performance de cuatro modelos de aprendizaje automático basados en regresión logística (LR), Bosques al azar (RF), Máquinas de soporte vectorial (SVM), y aumento de gradiente extremo (XGBoost), para

*Correspondence:
Syed Moazzam Nizami
moazzam.nizami@uoh.edu.pk
Qun ou Jiang
jiangqo@bjfu.edu.cn
Full list of author information is available at the end of the article

determinar la susceptibilidad en la predicción del fuego, y (iii) desarrolla un mapa de vulnerabilidad explícita para la región de Malakand, usando datos de ocurrencia de fuegos desde 2001 hasta 2023.

Métodos El modelado espacial fue conducido usando los productos de MODIS (FIRMS y MCD64A1), junto con variables climáticas, topográficas, vegetacionales, y de actividades humanas. Las relaciones de importancia en la predicción e inter-variables fueron analizadas antes de la puesta en marcha del modelo y de la evaluación.

Resultados El NDVI, la temperatura superficial, y la densidad de la población emergieron como los predictores claves. Dentro de los modelos probados, el RF y el XGBoost Superaron a los otros, con el RF alcanzando el 82% de exactitud y AUC 95,2. El análisis de multi-colinearidad confirmó la independencia de las variables (tolerancia > 0.1; VIF < 10).

Conclusiones El mapa resultante de la vulnerabilidad a los incendios delineó diferentes patrones espaciales, identificando el 6,88% del área de estudios como altamente susceptible a los incendios. Estos resultados proveen de explicaciones detalladas sobre los objetivos de prevención de incendios y manejo forestal sostenible en ecosistemas subtropicales del Himalaya

Introduction

Forest fires are an escalating global concern, causing substantial ecological and economic damage (Mehmood et al. 2024a; Latha 2022). Their rising frequency and intensity, driven by both climatic variability and anthropogenic pressures, underscore the urgent need for predictive fire risk assessments and effective management strategies (Baqer Rasooli and Bonyad, 2019; Ghorbanzadeh et al. 2019). Key ignition factors include topographic and meteorological conditions, vegetation characteristics, and socio-political drivers such as land use practices and policy responses (Pourtaghi et al. 2015; Ferreira et al. 2023). Climate change has amplified fire susceptibility in mountain ecosystems by increasing drought frequency and altering fire regimes, as demonstrated in recent studies from temperate regions (Dupire et al. 2019; Saidi et al. 2021; Shahzad et al. 2024)

The Malakand region, situated in the northwestern mountainous terrain of Pakistan, faces distinct forest fire challenges shaped by its unique topography, vegetation, and socio-economic dynamics. The area is characterized by steep slopes, dense chir pine forests (*Pinus roxburghii*), and a climate prone to extended dry periods, significantly increasing fire susceptibility. The region's forests play a critical ecological and economic role, acting as vital watersheds and providing livelihoods for local communities. However, increasing anthropogenic pressures such as grazing, agricultural expansion, and the traditional practice of setting fires to encourage new grass growth exacerbate the risk (Aslam et al. 2022). Coupled with limited firefighting infrastructure and inadequate forest management policies, these factors make the Malakand region particularly vulnerable to frequent and devastating forest fires. Moreover, socio-political challenges, including lack of awareness, weak institutional responses, and insufficient resource allocation, further complicate fire prevention and mitigation efforts (Pandey et al. 2023). Despite the increasing frequency of fire events in subtropical pine forests, most fire susceptibility studies have concentrated on temperate or Mediterranean ecosystems. In Pakistan's *Pinus roxburghii*-dominated forests, the interplay between anthropogenic pressures and biophysical drivers remains poorly quantified (Mehmood et al. 2025a, b). Furthermore, few studies have systematically evaluated the relative performance of multiple machine learning algorithms using harmonized environmental, climatic, and human influence data across fire-prone mountainous terrain (Hossen et al. 2024). This study addresses this gap by integrating multi-source geospatial datasets with model comparison to generate high-resolution fire vulnerability maps in the subtropical Himalayas.

Machine learning methods have gained prominence for their ability to analyze complex patterns and large datasets (Anees et al. 2024b). For instance, Chen et al. (2023) employed Random Forest (RF) and Support Vector Machine (SVM) models to predict forest fire susceptibility based on environmental variables, including vegetation indices, meteorological data, and historical fire occurrences. These machine-learning approaches handle non-linear relationships and integrate diverse datasets, making them highly effective for hazard prediction (Zhang et al. 2024). From a fire ecology standpoint, Chir Pine ecosystems are highly susceptible to fuel accumulation from needle litter, dense canopies, and persistent undergrowth, particularly under prolonged drought and unmanaged grazing. In wildland–urban interface (WUI) zones, where forests meet human settlements, ignition risks are further amplified by land use conflicts and resource pressures (Luo et al. 2025). Accurately modeling fire susceptibility in these settings necessitates algorithms that can capture complex, nonlinear interactions. Ensemble methods such as RF and XGBoost are well-suited for ecological applications due to their resistance to overfitting, capacity to model variable interactions, and provision of interpretable variable importance for management decisions (Mehmood et al. 2025a, b).

Despite the effectiveness of machine learning, traditional methods such as the Analytical Hierarchy Process (AHP) remain widely used due to their interpretability and ability to incorporate expert judgment. For example, Lamat et al. (2021) utilized AHP to generate risk maps, considering factors such as population density, LULC (land use/land cover), and meteorological conditions, categorizing areas into four risk levels: very high, high, moderate, and low. In another study, Das et al. (2023) integrated AHP with statistical modeling approaches like Frequency Ratio (FR) and Logistic Regression (LR) to assess forest fire susceptibility using topographic and climatic factors. Similarly, Pramanick et al. (2023) relied on GIS and AHP to delineate fire risk zones based on physical and socioeconomic variables.

Employing Geospatial Information Systems (GIS), simulation models, and remote sensing (RS) analysis provides robust tools for predicting forest fires and analyzing data from high-risk areas (Calkin et al. 2014). During vulnerability assessments, environmental factors such as altitude, precipitation, forest type, and terrain, alongside socioeconomic factors like population density and proximity to roads, play critical roles (Tariq et al. 2022). Chir pine forests (*Pinus roxburghii*), which dominate the subtropical pine zones, are particularly susceptible to human-induced fires due to slash-and-burn practices, grazing, and forest resource collection (Naskar et al. 2022). These activities, compounded by dry conditions, increased litter fall, and farmers' need for new grass growth, contribute to recurring fire incidents every 3 to 4 years (Singh et al. 2023). Additionally, canopy closure rates and vegetation density are critical determinants of fire risk, as lower canopy closure reduces risk, while dense vegetation often aligns with high-risk zones (Ozenen Kavlak et al. 2021). Despite advancements in fire assessment methodologies, several gaps remain in understanding fire dynamics in subtropical pine forests. Specifically, machine learning applications have largely focused on temperate and tropical forests, leaving subtropical ecosystems underexplored. Furthermore, integrating diverse environmental, topographic, and anthropogenic factors in predictive models remains limited, while rigorous model performance comparisons across such complex ecosystems are rare.

This study is novel in three main aspects: (i) it provides the first machine learning-based comparative fire susceptibility assessment specific to the Chir Pine forests of northern Pakistan, (ii) it integrates a comprehensive suite of variables including vegetation indices, climate reanalysis, topography, and human disturbance layers at high spatial resolution, and (iii) it employs both individual and ensemble machine learning algorithms to explore spatial heterogeneity in fire risk. Accordingly, the study aims to (1) identify key drivers of fire risk in subtropical pine forests, (2) evaluate and compare the performance of LR, RF, SVM, and XGBoost models, and (3) develop a vulnerability map to inform spatially targeted fire prevention strategies.

Methodology of the study

Study area

The study was conducted in the subtropical Chir Pine (*Pinus roxburghii*) zone of the Malakand region, located in northern Pakistan between 35° 29′ 59.99″ N and 72° 00′ 00.00″ E. The region Lies within the eastern margins of the Hindukush Mountain range, with elevations ranging from approximately 700 to 7708 m. Chir pine forests, situated between 800 and 1700 m, dominate the mid-elevation belt across Swat, Upper Dir, Lower Dir, Buner, Shangla, and Malakand districts. These forests are characterized by steep slopes, fragmented stands, and prolonged dry seasons, making them highly susceptible to fire ignition and spread. The climate features a distinct dry period from late spring to early autumn, which aligns with the region's fire season. Traditional fire use for pasture renewal, unmanaged grazing, and accumulated Surface Litter contribute to a recurring fire regime every 3–4 years. Approximately 10% (1.51 million hectares) of Pakistan's forest cover in this zone comprises subtropical Chir Pine forests, where fire incidents are frequently reported (Fig. 1).

Study design: data sources and preprocessing

The study utilized extensive fire data collected from January 2001 to December 2023, excluding the most recent decade. The MODIS fire product from the Fire Information for Resource Management System (FIRMS) was used to obtain active fire detections from NASA's Aqua and Terra satellites (https://firms.modaps.eosdis.nasa.gov) (Shahzad et al. 2024). The MCD64A1 Version 6 burned area product, a monthly global 500-m grid dataset, was integrated for spatial analysis. This dataset combines MODIS active fire observations with burned area classifications at a resolution of 1 km, which helps detect thermal anomalies and localize fires (Giglio et al. 2018; Lai et al. 2023; Li et al. 2025). Fire events with a reporting confidence level > 50% were aggregated to a 1 × 1 km grid system, with each cell coded as "1" for fire presence and "0" for absence. The aggregated fire data were validated against the MCD64A1 burned area product to ensure fire detection and delineation consistency. The dataset was split into 70% for training and 30% for testing, with

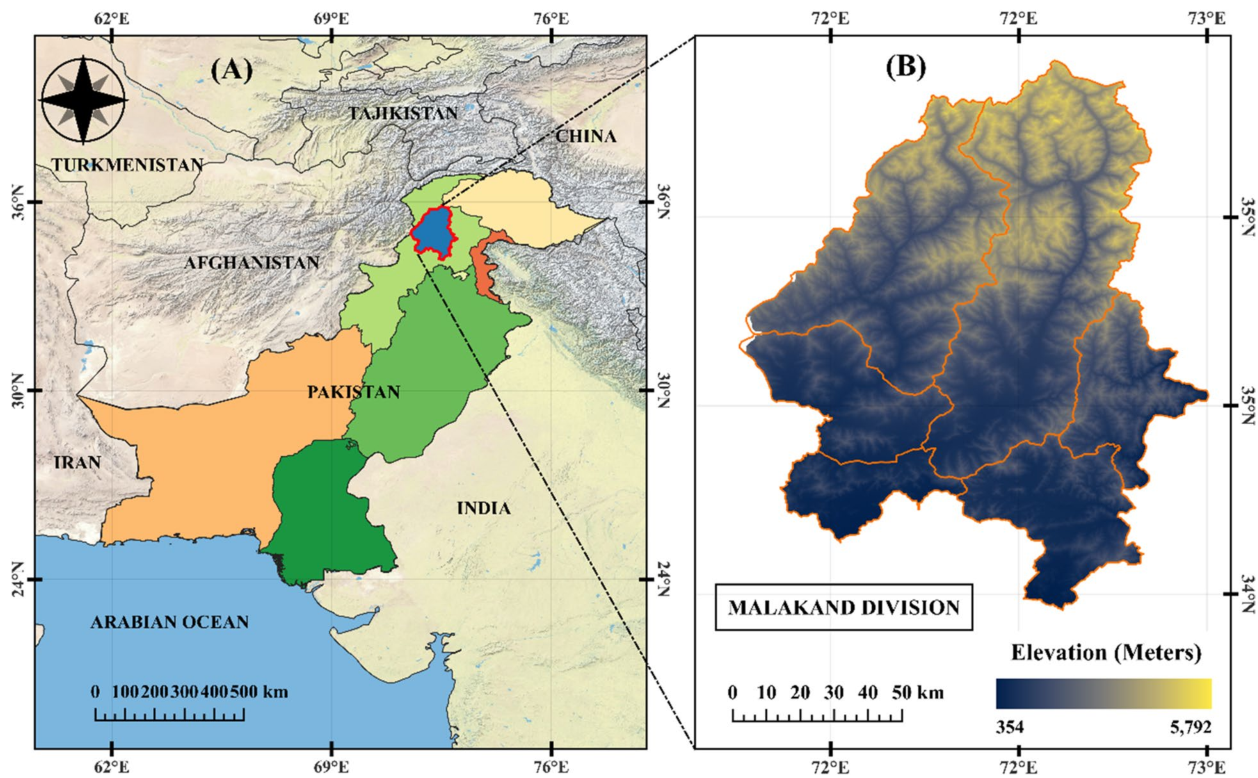

Fig. 1 Map of the study area. Panel **A** shows the geographical location of the Malakand Division within Pakistan alongside neighboring countries. Panel **B** provides a detailed elevation map of the Malakand Division, highlighting its topographic features and district boundaries

stratified sampling to maintain proportional fire and non-fire class representation. To improve model robustness and avoid overfitting, fivefold cross-validation was applied within the training set to tune model parameters and assess predictive stability. The held out 30% was reserved exclusively for final model performance evaluation.

Land cover information was derived from a 10-m resolution raster for 2020, obtained from the WorldCover project of the European Space Agency (ESA), based on Sentinel-1 and Sentinel-2 data (Sun et al. 2008). For topographic variables, slope, aspect, hill shade, and TWI were computed from the SRTM Digital Elevation Model (DEM) dataset using the ArcGIS Spatial Analyst extension. These topographic layers were resampled to 1×1 km resolution using nearest-neighbor interpolation to align with the fire and climatic datasets. All topographic variables were normalized to ensure compatibility across datasets. Monthly climatic data, including skin temperature, wind speed, and precipitation, were obtained from the ERA5 climate reanalysis dataset published by the Copernicus Climate Change Service (Zafar et al. 2023). Wind speed and skin temperature were selected over air temperature and solar radiation based on their direct ecological relevance to fire behavior. Wind speed is a well-established driver of fire spread dynamics, particularly in mountainous terrain where wind funnels through valleys and ridges. Skin temperature, derived from Surface thermal emissions, more accurately reflects near-ground heating and fuel dryness than ambient air temperature in remote sensing applications. Preliminary correlation analyses further confirmed that solar radiation exhibited low variability and poor association with historical fire detections and was therefore excluded to reduce feature redundancy. The climatic data were reprojected to match the spatial reference system of the study area and resampled to 1×1 km resolution using bicubic interpolation. The resulting layers were clipped to the study area for consistency with other datasets.

Anthropogenic variables such as population density were derived from the WorldPop dataset (Kolanek et al. 2021; Phelps and Beverly, 2022). Population density data were resampled to the 1×1 km grid system using bilinear interpolation. Proximity to roads, built-up areas, and rivers was calculated using OpenStreetMap data (Ehrig-Page 2020) and the Euclidean distance tool in GIS. Vegetation indices, such as the Normalized Difference Vegetation Index (NDVI), were extracted from the MOD13Q1 dataset, which provides data at a 500-m resolution (Table 1). Quality flags (QA) were applied to

Table 1 Independent variables with sources, units, and resolutions categorized by topographic, climatic, anthropogenic, and vegetation indices

Category	Independent variable	Source	Unit	Resolution
Topographic	Slope	https://earthexplorer.usgs.gov	m	30 m
	Aspect			
	Elevation			
	Hill shade			
	TWI			
Climatic	Skin temperature	ERA 5 climate reanalysis data	°C	9 km
	Wind speed		m/s	
	Precipitation		mm	
Anthropogenic	Population	https://hub.worldpop.org	count	1 km
	Proximity to road	https://www.openstreetmap.org	m	1:2,372,791
	Proximity to urban			
	Proximity to river			
Vegetation index	NDVI	MODIS/061/MOD13Q1	nm	250
	NMDI			

mask low-quality or cloud-covered pixels. The data were then clipped to the study area boundaries and temporally averaged to generate monthly composites. Missing values caused by cloud coverage were interpolated using nearest-neighbor resampling to ensure data completeness.

Analysis of multicollinearity and correlation of variables

The multicollinearity analysis was conducted using the Variance Inflation Factor (VIF) and tolerance (TOL) levels, commonly used to predict the relationships between causal variables. Multicollinearity is typically detected when the tolerance level (TOL) is less than 0.1, and the VIF is greater than 10, as per the commonly used rule of thumb (McCandless et al. 2020; Wang et al. 2019). These metrics were calculated using the following equations (Eqs. 1 and 2) (Khasawneh et al. 2024; Anees et al. 2024a)

$$TOL = 1 - \mathrm{R}^2 \tag{1}$$

$$VIF = \frac{1}{1-\mathrm{R}^2} = \frac{1}{TOL} \tag{2}$$

where R^2 denotes the coefficient of determination. In addition to the VIF and tolerance analysis, a Spearman Rank Correlation Heat map was employed further to examine the relationships between variables (Fig. 2).

Random Forest (RF)

Random Forest (RF) is a widely used ensemble algorithm based on decision trees, known for its robustness to overfitting and ability to model complex variable interactions (Breiman 2001). In this study, RF was applied to assess the contribution of environmental, topographic, and anthropogenic variables to fire occurrence. The number of trees and the number of features considered at each split were tuned using cross-validated grid search (Wai et al. 2022). The RF algorithm has a high level of predictability and is resistant to outliers, though it does have uncertainty. Besides, it proves efficient with abundant predictive accuracy, especially in forest fire prediction (Su et al. 2020). The quantity of random features and trees determines the predictive efficacy of the model (Jalal et al. 2022; Luo et al. 2024; Badshah et al. 2024). The RF model is one of the quickest machine-learning processes, capable of considering magnitude and scale with unprecedented explanation accuracies. However, it is sensitive to the problem of overfitting in practice (Arnaiz-González et al. 2016; Anees et al. 2024a).

Logistic Regression (LG)

Logistic Regression (LR) was employed to estimate the probability of vegetation fire occurrence as a function of environmental, topographic, and anthropogenic predictors. This method is widely used in binary classification problems due to its transparency, statistical interpretability, and suitability for modeling presence-absence data (Chas-Amil et al. 2012). Its effectiveness in fire prediction across diverse landscapes has been demonstrated in previous studies (Mallinis et al. 2019; Mohajane et al. 2021; Reyes-Bueno and Loján-Córdova, 2022). The mathematical formulation of the logistic function and variable interpretation is provided in Supplementary S1.

eXtreme Gradient Boosting (XGBoost)

XGBoost is a gradient-boosting framework that sequentially improves weak learners to minimize prediction

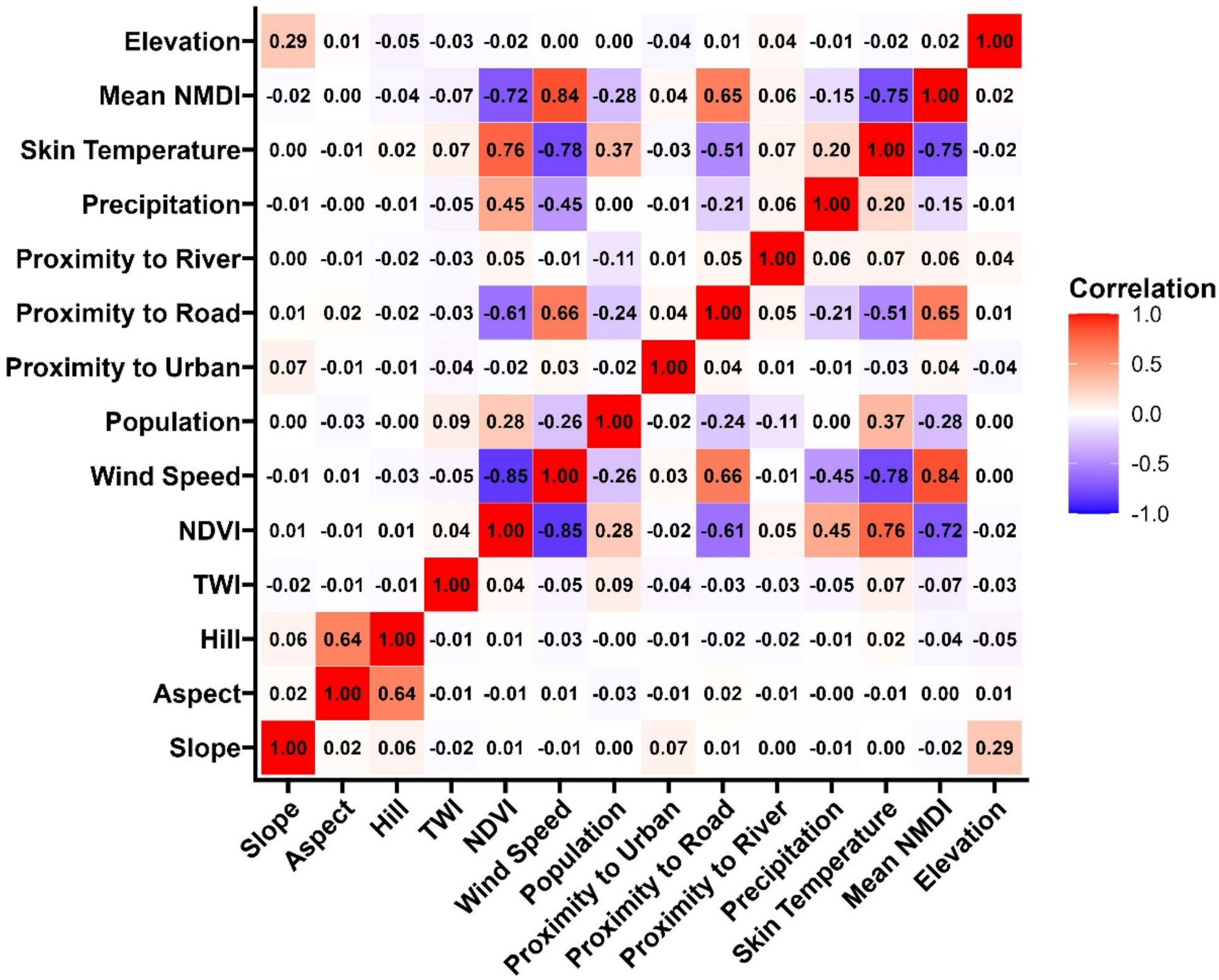

Fig. 2 Spearman rank correlation heatmap illustrating the relationships among environmental and anthropogenic variables

error. Its ability to handle missing data, penalize model complexity, and support regularization makes it suitable for non-linear fire susceptibility modeling (Chen and Guestrin, 2016; Tianqi and Carlos 2016). Key hyperparameters (learning rate, tree depth, number of estimators) were optimized via fivefold cross-validation. The base classifier in the XGBoost algorithm is CART. In XGBoost, each decision tree built during training is highly interrelated with the previous ones. The training and prediction results of each tree influence the construction of the next tree, allowing the model to correct errors and improve overall accuracy. Besides, it is a flexible algorithm in addition to solving regression and classification problems according to its targeted usage (Prasertsri and Littidej, 2020).

Support Vector Machine (SVM)

Support Vector Machine (SVM) was applied to classify fire and non-fire events by identifying optimal decision boundaries in high-dimensional feature space. A radial basis function (RBF) kernel was used to address potential non-linearity in the predictor variables. SVM is particularly effective for binary classification tasks involving large feature sets and sparse data, and has been successfully applied in earth observation studies, including forest fire prediction (Çakir et al. 2023). The model works by maximizing the margin between support vectors, allowing it to generalize well while minimizing overfitting (Rodriguez-Galiano et al. 2015; Xu et al. 2022). Hyperparameter tuning, including the penalty term (C) and kernel parameters (γ), was conducted using grid search with fivefold cross-validation to optimize classification accuracy (Kalantar et al. 2020; Liu et al. 2021; Zhang et al. 2023a, b).

Model comparison and statistical testing

To evaluate the performance of different machine learning models for forest fire susceptibility prediction, we compared LR, RF, SVM, and XGBoost using multiple evaluation metrics: Accuracy, AUC, Precision, Recall, and F1 Score. The metrics were calculated on test datasets following a 70:30 train-test split of the dataset. We performed bootstrapping to generate distributions of Accuracy values for each model to determine the statistical significance of performance differences between the models. Bootstrapping involved resampling the observed accuracy values 1000 times with replacement, creating robust distributions that reflect the inherent variability in model performance. Pairwise *t*-tests were then conducted to compare the bootstrapped Accuracy distributions for RF vs. LR, XGBoost vs. RF, and SVM vs. LR. The statistical tests evaluated whether the mean Accuracy differences between models were significantly different from zero, with results reported at a 95% confidence level (Sharma et al. 2022; Singha et al. 2024).

Results

Multicollinearity analysis of fire ignition factors

Table 2 Summarizes the multicollinearity analysis of different climate, anthropogenic, environmental, and topographic factors with their TOL being more than 0.1 and VIF being less than 10. This indicates that covariates among the factors of fire ignition based on the region can be utilized for forecasting future fire hazards.

The heatmap displays pairwise correlation coefficients, ranging from −1 (strong negative correlation) to 1 (strong positive correlation) (Fig. 2). Strong relationships, such as negative correlations between NDVI and Skin Temperature, and positive correlations among Population, Distance from Roads, and Urban Areas, highlight key interactions influencing fire risk. Weak correlations for variables like Elevation and Aspect indicate minimal direct associations. This analysis provides insights into the complex interplay of factors affecting fire susceptibility.

Table 2 Results of multicollinearity analysis

Variable	TOL	VIF
Elevation	0.89	1.11
Slope	0.90	1.11
Aspect	0.59	1.69
Hill shade	0.58	1.70
TWI	0.98	1.02
NDVI	0.22	4.44
Wind speed	0.12	7.96
Population	0.72	1.40
Distance from the road	0.98	1.01
Distance from the river	0.51	1.95
Distance from the urban	0.91	1.09
Precipitation	0.56	1.76
Skin temperature	0.26	3.82
NMDI	0.20	4.88

The heatmap reveals that some variables, such as NDVI and Skin Temperature, have notable negative correlations with other variables like Wind Speed and Precipitation, indicating that the other tends to decrease as one increases. This suggests that vegetation health, represented by NDVI, and moisture availability, represented by Precipitation, are inversely related to conditions like elevated Skin Temperature and increased Wind Speed (Mehmood et al. 2024c), both exacerbate fire risk. This relationship highlights the role of stressed vegetation and dry conditions in enhancing the likelihood of fire ignition. On the other hand, some variables, such as Elevation and Slope, show a very weak correlation with others, suggesting minimal direct interaction. While Elevation and Slope may not directly correlate with climatic or anthropogenic factors (Pan et al. 2023; Anees et al. 2024c; Muhammad et al. 2023; Gong et al. 2024), they indirectly influence fire spread by affecting fuel accumulation and fire behavior.

This detailed correlation analysis complements the multicollinearity analysis, offering a comprehensive understanding of how these environmental and anthropogenic factors interact in the context of fire risk. Together, these analyses provide critical insights into the relationships among the variables, helping to refine the predictive models by identifying and addressing potential multicollinearity issues and understanding the underlying correlations that may influence fire ignition and spread within the study area. For example, the negative correlation between NDVI and Skin Temperature reinforces the importance of vegetation health (Mehmood et al. 2024b) and moisture content in reducing fire risk, while positive relationships between Population and variables like Distance from the Road emphasize the impact of human activities on fire ignition potential.

Evaluation of machine learning model performance

The performance of four machine learning models Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost were assessed using five evaluation metrics: Accuracy, Area Under the Curve (AUC), Precision, Recall, and F1 Score. These metrics offer complementary insights into the models' classification performance and guide the selection of the most effective classifier. The LR model achieved an accuracy of 81.6%, indicating that approximately four out of five instances were correctly classified. The AUC score of 88.7 reflects moderate discriminatory ability between fire and non-fire cases. In comparison, all three other models

Table 3 Performance metrics of different machine learning models in classification tasks

Model types	Accuracy	AUC	Precision	Recall	F1 score
Logistic regression	81.6	88.7	84.7	76.9	80.6
Random forest	86.2	95.4	88.1	83.5	85.8
SVM	81.6	90.7	92.8	68.4	78.7
XGBoost	87.2	95.2	90.0	83.5	86.3

Random Forest (RF), Support Vector Machine (SVM), and XGBoost demonstrated higher performance across most evaluation metrics (Table 3). As shown in Fig. 3, the proximity of their ROC curves to the top-left corner indicates improved sensitivity and specificity relative to LR. The precision of the LR model was 84.7%, while the recall was lower at 76.9%, resulting in an F1 Score of 80.6%. These values Suggest that although LR performs reasonably well, it is less effective in capturing fire-prone instances compared to the ensemble classifiers. The RF model outperformed Logistic Regression on all major metrics. It achieved an accuracy of 86.2% and an AUC of 95.4, indicating a strong capacity to distinguish between classes. Precision and recall were 88.1 and 83.5%, respectively, yielding an F1 Score of 85.8%. These results reflect a robust and balanced model with higher classification reliability for fire susceptibility than LR. The superiority of RF over LR was statistically confirmed through bootstrapped accuracy comparisons (mean difference = 4.64%, $p < 0.005$; Table 4).

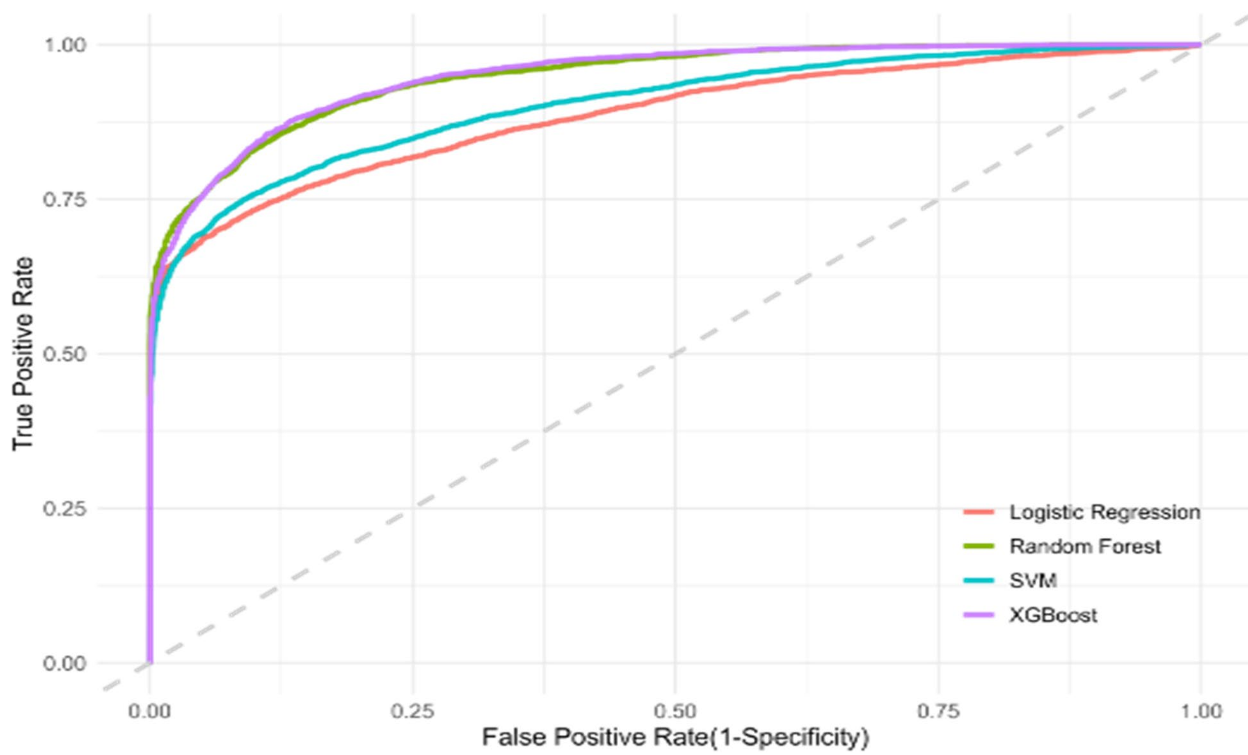

Fig. 3 Receiver operating characteristic (ROC) curves for Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost models in predicting forest fire ignition. The Area Under the Curve (AUC) scores for Logistic Regression = 88.7, SVM = 90.7, Random Forest = 95.4, and XGBoost = 95.2

Table 4 Pairwise statistical testing results

Comparison	(%)Mean difference	*t*-value	*p*-value	95% confidence interval	Sig:
Random Forest vs Logistic Regression	4.64	52.447	< 0.005	[4.47, 4.81]	Sig:
XGBoost vs Random Forest	0.97	11.287	< 0.005	[0.80, 1.14]	Sig:
SVM vs Logistic Regression	0.045	0.492	> 0.005	[− 0.14, 0.23]	Not Sig:

The ROC curves illustrate the trade-off between the True Positive Rate (Sensitivity) and the False Positive Rate (1 − Specificity) for each classifier (Fig. 3). A curve closer to the top-left corner reflects stronger performance in distinguishing between classes. Among the four models, Random Forest and XGBoost achieved the highest AUC values, indicating strong discriminatory power in predicting forest fire Susceptibility. The SVM model achieved an accuracy of 81.6%, which was equal to that of Logistic Regression. However, it demonstrated a higher AUC score of 90.7, suggesting better overall separability between fire and non-fire classes. The model produced the highest precision (92.8%) among all classifiers, indicating that when it predicted a fire-prone area, it was frequently correct. However, its recall value was the lowest (68.4%), indicating a tendency to miss a considerable number of actual fire cases. As a result, the F1 Score was 78.7%, reflecting an imbalance between sensitivity and specificity. These results suggest that SVM is particularly strong in minimizing false positives but less effective in identifying all positive instances.

In comparison, XGBoost achieved the highest overall classification performance. Its accuracy reached 87.2%, the highest among all models. The AUC score was 95.2 slightly lower than Random Forest but still indicating excellent class discrimination. The model produced a precision of 90.0% and a recall of 83.5%, resulting in the highest F1 Score of 86.6%. This combination of high sensitivity and precision positions XGBoost as the most balanced and reliable classifier for fire susceptibility prediction when both false positives and false negatives are of concern. Across all models, Random Forest and XGBoost consistently demonstrated the strongest performance based on accuracy, AUC, and F1 Score (Table 3). The superiority of these two ensemble classifiers was statistically validated using bootstrapped accuracy distributions and pairwise *t*-tests (Table 4). XGBoost showed a statistically significant improvement over Random Forest (mean difference = 0.97%, $p < 0.005$), while Random Forest significantly outperformed Logistic Regression (mean difference = 4.64%, $p < 0.005$). The performance difference between SVM and Logistic Regression was not statistically significant ($p > 0.05$).

While all models showed acceptable classification ability, Random Forest and XGBoost were the most suitable for deployment in similar fire risk modeling tasks due to their superior balance between precision and recall. XGBoost is advantageous for modeling complex, nonlinear interactions through iterative boosting, with built-in regularization to mitigate overfitting. This makes it well-suited for modeling interactions among climatic, environmental, and anthropogenic variables. In contrast, Random Forest remains an efficient and interpretable option, especially useful for large or noisy datasets. It provides stable predictions and clear variable importance scores, aiding ecological interpretation, particularly for fire risk drivers like NDVI and skin temperature. Although XGBoost slightly outperformed Random Forest in predictive metrics, both ensemble models offer robust, generalizable performance for forest fire vulnerability assessment.

Based on the evaluation metrics, RF and XGBoost demonstrated the strongest classification performance, particularly in terms of Accuracy, AUC, and F1 Score (Table 3). To statistically validate these differences, we conducted pairwise comparisons using bootstrapped accuracy distributions (1000 resamples) followed by paired *t*-tests. The results indicated that Random Forest significantly outperformed LR, with a mean accuracy difference of 4.64% ($t = 52.447$, $p < 2.2\mathrm{e}-16$, 95% CI [4.47, 4.81]). XGBoost also significantly outperformed RF, albeit with a smaller margin of 0.97% ($t = 11.287$, $p < 2.2\mathrm{e}-16$, 95% CI [0.80, 1.14]). The difference in accuracy between SVM and Logistic Regression was not statistically significant (mean difference = 0.045%, $t = 0.492$, $p = 0.6225$, 95% CI [−0.14, 0.23]) (Table 4). These statistical results confirm that both Random Forest and XGBoost provided significantly better classification accuracy than Logistic Regression, with XGBoost showing a marginal but statistically significant improvement over Random Forest. The findings support the selection of ensemble learning methods particularly XGBoost for forest fire susceptibility modeling under complex environmental conditions.

Comparative importance of fire risk triggers in the Malakand Division

Figure 4 presents the relative importance of predictor variables in determining fire ignition probability across the Malakand Division, based on the Random Forest model. NDVI emerged as the most significant factor, underscoring the critical role of vegetation health and fuel accumulation in influencing fire susceptibility in subtropical pine ecosystems (Mehmood et al. 2024d). The second most important variable was skin temperature, which serves as a proxy for surface heating and fuel dryness, both well-established contributors to fire ignition. Among the anthropogenic and climatic variables, population density, proximity to water bodies (km), and precipitation were ranked as influential. These variables likely represent patterns of human presence, fuel moisture variability, and ignition potential near accessible landscapes. Moderate importance was observed for wind speed and proximity to roads, indicating that while these factors contribute to ignition dynamics, their influence may be more relevant to fire spread or accessibility than ignition probability alone.

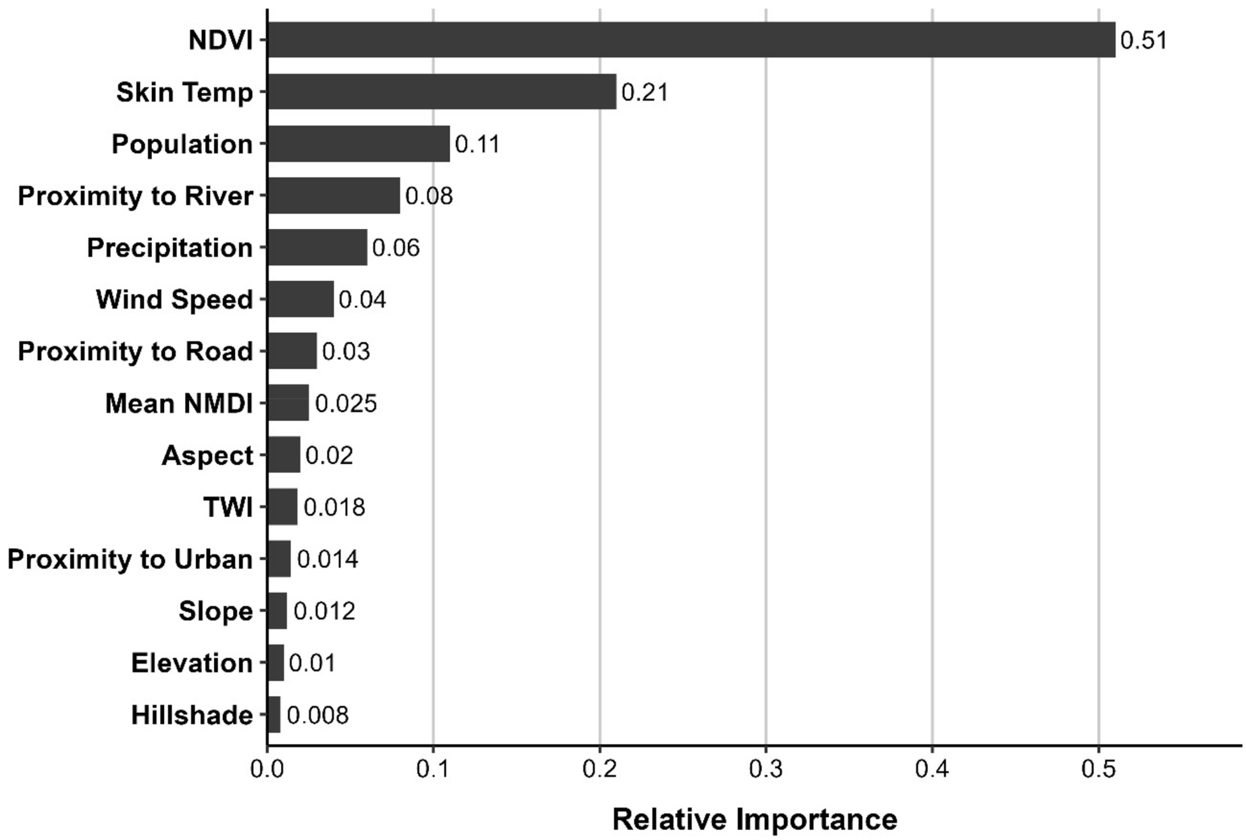

Fig. 4 Variable importance plot highlighting key factors in predicting fire ignition in the Malakand Division

Surprisingly, terrain variables such as elevation, slope, aspect, topographic wetness index (TWI), and hill shade exhibited low relative importance in the model. Although these factors are often associated with fire behavior and spread, their direct effect on fire ignition appears limited in this landscape. This suggests that ignition in the Chir Pine forests of Malakand is more strongly driven by surface vegetation conditions and human activity than by topographic variability possibly due to the prevalence of anthropogenic ignitions across elevations and slope gradients. These results highlight the combined influence of environmental stressors, land cover conditions, and anthropogenic pressures in driving ignition patterns. The low importance of terrain-related variables further emphasizes the need to prioritize vegetation monitoring and human activity regulation in fire risk reduction strategies.

Table 5 Distribution of fire risk categories by area in the Malakand Division

Class	Severity level	Area (km^2)	Percentage (%)
1	Very high risk	1011.409	6.88%
2	High risk	2518.155	17.12%
3	Intermediate risk	4835.196	32.88%
4	Low risk	2265.652	15.41%
5	Very low risk	4075.702	27.71%

Fire risk mapping results and critical analysis

The fire risk map for the Malakand Division reveals a heterogeneous spatial distribution of fire susceptibility. The study area was categorized into five fire risk levels: Very High, High, Intermediate, Low, and Very Low, based on the natural breaks (Jenks) classification applied to the fire susceptibility model outputs. This method optimally partitions values into classes with the least within-class variance, providing clearer distinctions between risk levels as shown in Table 4. According to Table 5, the largest share of the area (32.88%) falls within the Intermediate Risk category, followed by Very Low (27.71%), High (17.12%), Low (15.41%), and Very High Risk (6.88%). Although limited in extent, Very High-Risk areas represent critical ignition zones, typically characterized by a combination of dense vegetation (high NDVI), elevated skin temperature, and Limited moisture availability. These areas are most found along south-facing slopes between 800 and

1200 m and often border settled zones or degraded forest areas with recurring human disturbance, such as grazing, fuelwood harvesting, and agriculture.

High-Risk zones, while not as extreme, still warrant concern and cover 17.12% of the region. These areas often function as transitional zones between urbanized and forested landscapes, where vegetation remains abundant but is increasingly intersected by anthropogenic activities. In spatial terms, high-risk zones were observed across parts of Lower Dir, Shangla, and central Malakand, where both climatic stress and proximity to road networks increase the ignition probability. Susceptibility in the entire area of the Malakand division, as indicated in this fire risk map, is heterogeneous (Fig. 5). These are further categorized into Very High, High, Intermediate, Low, and Very Low according to the Fire Prediction Areas (Table 5).

Intermediate Risk zones, which account for the largest spatial extent, reflect regions with a mix of vegetation types, moderate climatic conditions, and varied topography. These areas are generally more stable but could shift to higher-risk categories under prolonged drought or intensified land use. Their presence across large parts of Upper Dir and eastern Buner emphasizes the need for Sustainable land management to prevent ecological degradation. Low and Very Low-Risk areas cover nearly 43% of the division and are largely situated in elevated, cooler, or sparsely vegetated zones, where ignition likelihood is naturally lower. However, these zones still require baseline monitoring and conservation efforts to maintain their current risk profile.

The spatial insights from this risk mapping are highly relevant for regional forest fire planning. We recommend that fire control resources be prioritized in northwestern Swat and southern Buner, particularly along 800–1200 m elevation zones with overlapping high NDVI and skin temperature values. These areas

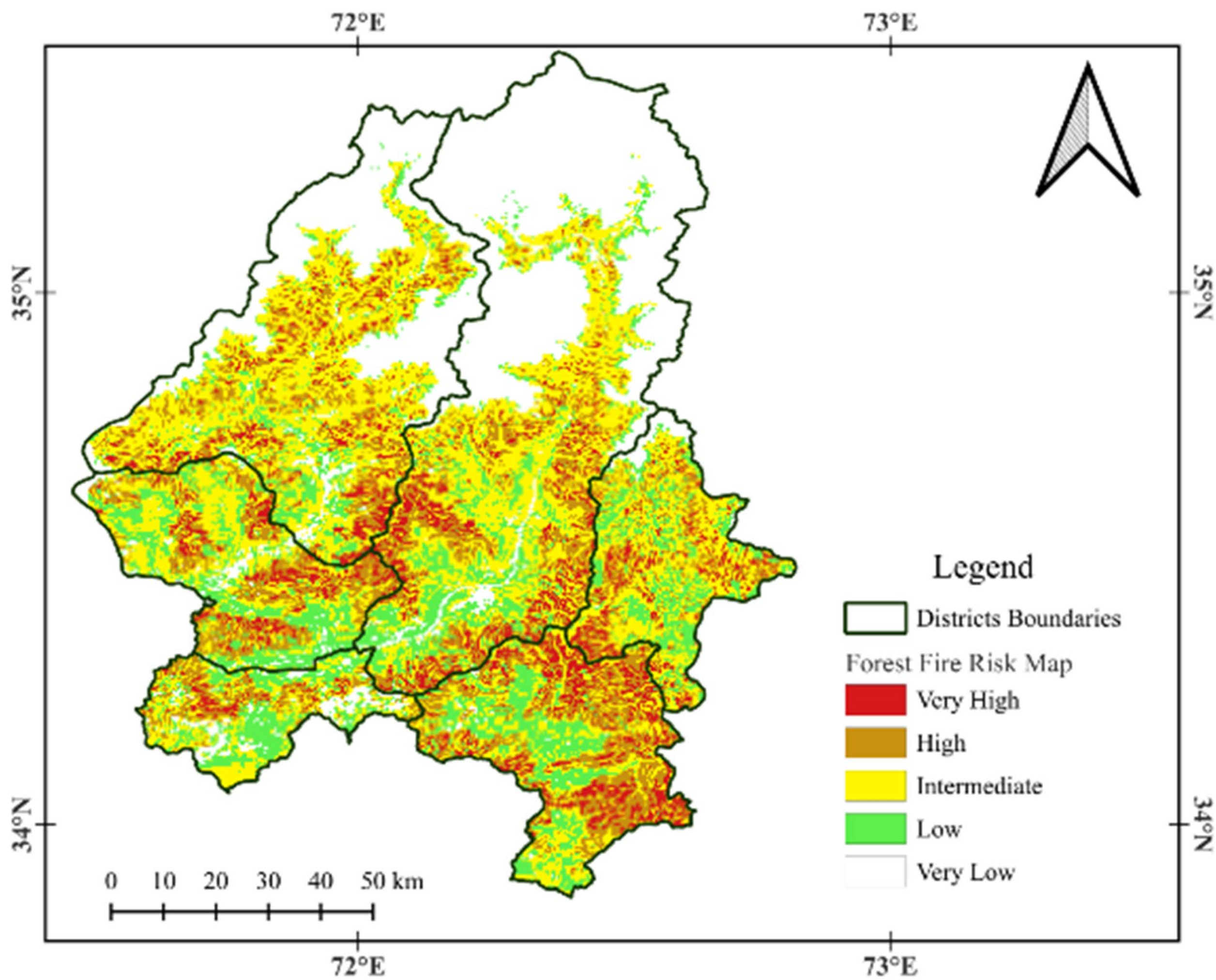

Fig. 5 Forest fire risk map of the Malakand Division, highlighting risk levels from very high to very low

are consistently associated with higher ignition probabilities. Furthermore, the risk mapping outputs can be integrated into early warning systems, seasonal fire preparedness plans, and community-based fire mitigation programs. This approach aligns with national forest monitoring initiatives and supports spatially targeted resource allocation for fire prevention, especially in fire-prone administrative units within the KP region.

Discussion

This study enhances understanding of fire ignition dynamics in the Malakand Division by integrating geospatial predictors with ensemble machine learning models. NDVI and skin temperature were identified as the most influential variables, underscoring the ecological significance of vegetation condition and surface energy in fire susceptibility modeling. In *Pinus roxburghii* forests, fluctuations in NDVI capture critical biophysical processes such as reductions in live biomass, understory desiccation, and the accumulation of resinous needle litter that directly increases fuel flammability (Mehmood et al. 2024e; Shobairi et al. 2022; Usoltsev et al. 2020, 2022). These forests, characterized by open canopies and coarse fuels, become highly fire-prone under conditions of vegetative stress and elevated surface temperature (Yathish et al. 2019).

Skin temperature reinforces these ignition conditions by intensifying evapotranspirative loss and reducing moisture retention in both live and dead fuels (Andreevich et al. 2020; Anees et al. 2022a; Jiao et al. 2024; Du et al. 2024). The interaction between declining NDVI and rising thermal stress explains the high ignition probabilities in densely vegetated, topographically exposed areas (Manzo-Delgado et al. 2009). These results are consistent with fire ecology studies integrating NDVI and land surface temperature (Abdollahi et al. 2018; Michael et al. 2021) but extend their utility to subtropical Himalayan systems.

Unlike studies in peri-urban or high-density landscapes (Clarke et al. 2019; Kolanek et al. 2021; Ricotta et al. 2018; Zhang et al. 2022), where anthropogenic drivers dominate, this study found that population density and road proximity while statistically valid predictors (VIF = 1.40 and 1.01, respectively) contributed less to fire ignition risk than environmental variables. This outcome reflects the dominance of landscape-driven ignition processes in the Malakand Division, where fuel structure, vegetation stress, and climatic stressors are primary determinants of fire risk, even in proximity to human activity.

Evaluation of machine learning models

The Superior performance of the RF and XGBoost models in predicting fire risk underscores the effectiveness of ensemble methods in handling complex, non-linear interactions among variables. The RF model achieved an accuracy of 86.2% and an AUC of 95.4, while XGBoost reached an accuracy of 87.2% and an AUC of 95.2, demonstrating their suitability for operational fire risk assessment. These results align with other studies comparing RF and XGBoost in environmental applications. For instance, Mehmood et al. (2024a) reported high AUC values (0.935 to 0.953) for both models in forest fire risk prediction, highlighting their robustness. Rabby et al. (2022) also found XGBoost to outperform RF in landslide susceptibility mapping, emphasizing XGBoost's superior handling of spatial data. Similarly, Eskandari et al. (2020) confirmed RF's reliability in mapping fire hazards, with an AUC of 0.855. Further research supports these findings, such as Dong et al. (2022), who found XGBoost slightly better than RF in wildfire risk prediction. Additionally, Liu et al. (2021) confirmed that both models could accurately predict net ecosystem exchanges across various biomes, reinforcing their utility in environmental modeling. On the other hand, the Support Vector Machine (SVM) model, while achieving a reasonable accuracy of 81.6%, showed a lower recall of 68.4%, indicating potential limitations in capturing certain positive instances of fire occurrences. This result aligns with studies that highlight SVM's challenges in handling complex, high-dimensional data, particularly when non-linear relationships are involved (Zhang et al. 2018). These limitations should be considered when selecting SVM for similar environmental modeling tasks in future research.

Implications for fire management and policy

The heterogeneous distribution of fire risk across the Malakand Division indicates the need for *targeted fire management strategies* based on spatial risk gradients. The fire risk map revealed that 6.88% of the area falls under the Very High-Risk category, while 17.12% is classified as High Risk. These zones, often associated with *dense vegetation*, *elevated surface temperatures*, and *reduced moisture*, require immediate attention for fire prevention and mitigation. Concentrated clusters were identified in regions such as *northwestern Swat*, *southern Buner*, and *mid-elevation belts (800–1200 m)* areas where high NDVI and skin temperature values co-occur. Fire control resources should be prioritized in these high-risk areas through measures such as *fuel load reduction*, *canopy management*, and *early detection systems*. This spatially informed approach aligns with recommendations from Cai et al. (2024), who emphasize localized intervention strategies. In Liangshan Yi Autonomous Prefecture, China, Zhang et al. (2023a, b) demonstrated that regionalizing fire risks based on local environmental conditions significantly improved fire management outcomes.

Similarly, Laha et al. (2021) applied multi-factor fire risk mapping in Sikkim, India, and achieved more accurate identification of priority areas for fire prevention. Beyond the highest risk zones, the broader distribution of Intermediate and Low-Risk areas covering 32.88 and 15.41% of the region respectively suggests that fire risk is not confined to isolated hotspots. These areas represent transitional landscapes that, if unmanaged, may evolve into higher-risk zones under climatic stress or anthropogenic pressure. Effective strategies in these regions include regulated grazing, sustainable harvesting, and moisture conservation practices, aimed at maintaining ecological resilience. This is consistent with integrated fire management approaches proposed by Khan et al. (2024) and Yan et al. (2023), who emphasize maintaining landscape stability as a preventive tool. To operationalize these insights, the fire risk outputs generated in this study can be integrated into regional forest planning tools and early warning systems, enhancing preparedness and resource allocation. Spatial overlays of predicted fire risk with infrastructure, land use, and topographic vulnerability layers can support data-driven decision-making in future fire policy design and reserve management (Li et al. 2024).

Limitations

This study provides valuable insights into forest fire risk modeling in the Malakand Division, but several limitations warrant consideration. The reliance on remotely sensed datasets, while advantageous for capturing broad spatial and temporal patterns, introduces uncertainty related to sensor resolution, accuracy, and data quality (Fang et al. 2025). For example, MODIS NDVI, with its 250-m spatial resolution, may not detect fine-scale variations in vegetation structure and understory fuel conditions (Anees et al. 2022b), potentially reducing the spatial precision of fire risk predictions (Mehmood et al. 2024a; Guo et al. 2024). Similarly, the 8-day compositing interval may overlook rapid changes in vegetation stress or fire spread, limiting the temporal responsiveness of the model. Another limitation is the geographic scope of the study. While the Malakand Division provides a representative subtropical pine ecosystem, the findings may not be fully generalizable to other forest types, climatic zones, or socio-ecological settings. Models trained on this region may underperform when transferred to areas with different vegetation composition, elevation gradients, or anthropogenic pressures. Although the RF and XGBoost models demonstrated strong predictive performance, ensemble methods are inherently prone to overfitting when applied to region-specific, high-dimensional datasets without appropriate regularization (Shekar et al. 2025). Additional ecological variables such as soil moisture, soil texture, and organic content were excluded due to data availability constraints, possibly limiting the models' ability to fully represent fire-enabling conditions. This study did not include field-based ground validation or local fire incident records due to logistical and data access limitations. Future studies should incorporate ground truthing using forest department fire logs, GPS-referenced burn sites, or community-reported fire incidents to validate and calibrate remote sensing-based fire occurrence layers more accurately. Furthermore, the current model does not consider fire suppression infrastructure, such as the presence of fire lines, lookout posts, forest access roads, or firefighting units, which could significantly influence ignition containment and fire spread. Integrating spatial layers of suppression capacity into future fire risk models would enhance their operational utility for disaster preparedness and resource allocation. Lastly, like most historical models, the analysis assumes stationary climate-fire relationships, which may not hold under accelerated climate change scenarios. Future work should explore dynamic modeling approaches that incorporate seasonal forecasts, near-real-time indicators, and machine learning ensembles trained on multi-scenario climate projections to improve the resilience of fire risk assessments under future uncertainty (Akram et al. 2022; Hussain et al. 2024a, b; Haider et al. 2017; Jallat et al. 2021; Khan et al. 2020).

Conclusion

This study demonstrates the efficacy of machine learning models in assessing forest fire vulnerability in subtropical pine ecosystems, particularly through the integration of environmental, topographic, climatic, and anthropogenic variables. The following key findings summarize the core outcomes:

- RF and XGBoost were the most effective models, achieving the highest predictive accuracy and AUC scores for ignition risk classification.
- NDVI and skin temperature were the dominant predictors, reflecting the critical role of vegetation health and thermal stress in fire susceptibility across *Pinus roxburghii* landscapes.
- The fire vulnerability map revealed distinct high-risk zones, especially in regions with dense canopy cover and elevated thermal loads, supporting the need for spatially targeted fire mitigation strategies.

These insights offer valuable guidance for fire managers and policymakers. Priority should be given to high-risk zones identified in this study for interventions such as controlled fuel load reduction, early warning systems, and remote sensing-based monitoring. In densely inhabited or degraded forest areas, community awareness campaigns and investments in firebreak infrastructure are essential to minimize anthropogenic ignition triggers. To enhance model precision, future work should consider incorporating higher-resolution satellite products (e.g., Sentinel-2) and ground-based fire incident records for validation. While deep learning approaches such as convolutional neural networks (CNNs) offer promising capabilities in capturing spatial complexity, their applicability in subtropical forest fire modeling warrants empirical evaluation. Further research should explore their potential alongside ensemble methods, particularly for developing adaptive and transferable fire risk frameworks under changing climatic conditions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42408-025-00408-w.

Supplementary Material 1.

Acknowledgements

This research is conducted by the Department of Forestry and wildlife Management, The University of Haripur. Pakistan and School of Soil and Water conservation, Beijing Forestry University China. We extend our thanks to all those faculty members and officials of the department concerned for granting us access to the location and providing valuable information for the study.

Authors' contributions

Conceptualization, MN, JQ; Data curation, SM, KM, FS; Formal analysis, SM, MH, KM, SAA, Investigation, SAA, WRK; Methodology, SM, KM; Project administration: MH, Software: KM, SAA, JQ; Supervision, MN, JQ; Validation, SAA, WRK, FS, Visualization, SM, MH, SAA; Writing—original draft, SM, KM; Writing—review & editing, MN.

Funding

This work was financially supported by the National Natural Science Foundation of China (NSFC) grant no:42371291.

Data availability

Data will be made available on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Forestry and Wildlife, Faculty of Physical & Applied Sciences, The University of Haripur, Hattar Road, Haripur, Pakistan. [2]Institute of Forest Sciences, University of Swat, Mingora Charbagh, Khyber-Pakhtunkhwa, Pakistan. [3]School of Soil & Water Conservation, Beijing Forestry University, Beijing 100083, China. [4]College of Forestry, Beijing Forestry University, Beijing, China. [5]Department of Forestry, The University of Agriculture, Dera Ismail Khan 29050, Pakistan. [6]Precision Forestry Key Laboratory of Beijing, Beijing Forestry University, Beijing 100083, People's Republic of China. [7]Department of Forestry Science and Biodiversity, Faculty of Forestry and Environment, Universiti Putra Malaysia, Serdang 43400, Malaysia.

Received: 8 February 2025 Accepted: 25 August 2025
Published online: 11 November 2025

References

Abdollahi, M., T. Islam, A. Gupta, and Q. K. Hassan. 2018. An advanced forest fire danger forecasting system: Integration of remote sensing and historical sources of ignition data. *Remote Sensing* 10 (6): 923.

Akram, M., U. Hayat, J. Shi, and S. A. Anees. 2022. Association of the female flight ability of Asian spongy moths (*Lymantria dispar asiatica*) with locality, age and mating: A case study from China. *Forests* 13 (8): 1158. https://doi.org/10.3390/f13081158.

Andreevich, U. V., S. S. O. Reza, T. I. Stepanovich, A. Amirhossein, Z. Meng, S. A. Anees, and C. V. Petrovich. 2020. Are there differences in the response of natural stand and plantation biomass to changes in temperature and precipitation? A case for two-needled pines in Eurasia. *Journal of Resources and Ecology* 11 (4): 331. https://doi.org/10.5814/j.issn.1674-764x.2020.04.001.

Anees, S. A., K. Mehmood, A. Rehman, N. U. Rehman, S. Muhammad, F. Shahzad, K. Hussain, M. Luo, A. A. Alarfaj, S. A. Alharbi, and W. R. Khan. 2024b. Unveiling fractional vegetation cover dynamics: A spatiotemporal analysis using MODIS NDVI and machine learning. *Environmental and Sustainability Indicators*. https://doi.org/10.1016/j.indic.2024.100485.

Anees, S. A., K. Mehmood, W. R. Khan, M. Sajjad, T. A. Alahmadi, S. A. Alharbi, and M. Luo. 2024a. Integration of machine learning and remote sensing for above ground biomass estimation through Landsat-9 and field data in temperate forests of the Himalayan region. *Ecological Informatics*. https://doi.org/10.1016/j.ecoinf.2024.102732.

Anees, S. A., X. Yang, and K. Mehmood. 2024c. The stoichiometric characteristics and the relationship with hydraulic and morphological traits of the Faxon fir in the subalpine coniferous forest of Southwest China. *Ecological Indicators* 159 : 111636. https://doi.org/10.1016/j.ecolind.2024.111636.

Anees, S. A., X. Zhang, K. A. Khan, M. Abbas, H. A. Ghramh, and Z. Ahmad. 2022a. Estimation of fractional vegetation cover dynamics and its drivers based on multi-sensor data in Dera Ismail Khan, Pakistan. *Journal of King Saud University - Science* 34 (6) : 102217. https://doi.org/10.1016/j.jksus.2022.102217.

Anees, S. A., X. Zhang, M. Shakeel, M. A. Al-Kahtani, K. A. Khan, M. Akram, and H. A. Ghramh. 2022b. Estimation of fractional vegetation cover dynamics based on satellite remote sensing in pakistan: A comprehensive study on the FVC and its drivers. *Journal of King Saud University-Science* 34 (3) : 101848. https://doi.org/10.1016/j.jksus.2022.101848.

Arnaiz-González, Á., J. F. Díez-Pastor, C. García-Osorio, and J. J. Rodríguez. 2016. Random feature weights for regression trees. *Progress in Artificial Intelligence* 5 (2): 91–103.

Aslam, M. S., P. Huanxue, S. Sohail, M. T. Majeed, S. U. Rahman, and S. A. Anees. 2022. Assessment of major food crops production-based environmental efficiency in China, India, and Pakistan. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-021-16161-x.

Badshah, M. T., K. Hussain, A. U. Rehman, K. Mehmood, B. Muhammad, R. Wiarta, R. F. Silamon, M. A. Khan, and J. Meng. 2024. The role of random forest and Markov chain models in understanding metropolitan urban growth trajectory. *Frontiers in Forests and Global Change* 7:1345047.

Baqer Rasooli, S., and A. E. Bonyad. 2019. Evaluating the efficiency of the Dong model in determining fire vulnerability in Iran's Zagros forests. *Journal of Forestry Research* 30 (4): 1447–1458.

Breiman, L. 2001. Random forests. *Machine Learning* 45:5–32.

Cai, G., X. Zheng, W. Gao, and J. Guo. 2024. Self-extinction characteristics of fire extinguishing induced by nitrogen injection rescue in an enclosed urban utility tunnel. *Case Studies in Thermal Engineering* 59 : 104478. https://doi.org/10.1016/j.csite.2024.104478.

Çakir, M., M. Yilmaz, M. A. Oral, H. Ö. Kazanci, and O. Oral. 2023. Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture. *Journal of King Saud University-Science* 35 (6) : 102754.

Calkin, D. E., J. D. Cohen, M. A. Finney, and M. P. Thompson. 2014. How risk management can prevent future wildfire disasters in the

wildland-urban interface. *Proceedings of the National Academy of Sciences.* 111 (2): 746–751.
Chas-Amil, M. L., E. García-Martínez, and J. Touza. 2012. Fire risk at the wildland-urban interface: A case study of a Galician county. *WIT Transactions on Ecology and the Environment* 158:177–188.
Chen, T. and Guestrin, C. 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785–794).
Chen, Y., X. Zhang, W. Li, and H. Zhao. 2023. Forest fire susceptibility mapping using random forest and support vector machine models: A case study in the northeastern region. *Journal of Environmental Management* 315 : 110992. https://doi.org/10.1016/j.jenvman.2023.110992.
Clarke, H., R. Gibson, B. Cirulis, R. A. Bradstock, and T. D. Penman. 2019. Developing and testing models of the drivers of anthropogenic and lightning-caused wildfire ignitions in south-eastern Australia. *Journal of Environmental Management* 235:34–41.
Das, J., S. Mahato, P. K. Joshi, and Y. A. Liou. 2023. Forest fire susceptibility zonation in eastern India using statistical and weighted modelling approaches. *Remote Sensing* 15 (5): 1340.
Dong, H., H. Wu, P. Sun, and Y. Ding. 2022. Wildfire prediction model based on spatial and temporal characteristics: A case study of a wildfire in Portugal's Montesinho Natural Park. *Sustainability* 14 (16): 10107.
Du, C., X. Bai, Y. Li, Q. Tan, C. Zhao, G. Luo, J. Wang, L. Wu, C. Li, J. Li, and Y. Xie. 2024. The restoration of karst rocky desertification has enhanced the carbon sequestration capacity of the ecosystem in southern China. *Global and Planetary Change* 243 : 104602.
Dupire, S., T. Curt, S. Bigot, and T. Fréjaville. 2019. Vulnerability of forest ecosystems to fire in the French Alps. *European Journal of Forest Research* 138 (5): 813–830.
Ehrig-Page, J. C. 2020. Evaluating methods for downloading OpenStreetMap data. *Cartographic Perspectives* 95:42–49.
Eskandari, S., Amiri, M., Sādhasivam, N., and Pourghasemi, H. R., 2020. Comparison of new individual and hybrid machine learning algorithms for modeling and mapping fire hazard: a supplementary analysis of fire hazard in different counties of Golestan Province in Iran. *Natural Hazards.* 104(1). https://doi.org/10.1007/s11069-020-04169-4
Fang, C., K. Song, Z. Yan, and G. Liu. 2025. Monitoring phycocyanin in global inland waters by remote sensing: Progress and future developments. *Water Research* 275 : 123176. https://doi.org/10.1016/j.watres.2025.123176.
Ferreira, I. J., W. A. Campanharo, M. L. Barbosa, S. S. D. Silva, G. Selaya, L. E. Aragão, and L. O. Anderson. 2023. Assessment of fire hazard in southwestern Amazon. *Frontiers in Forests and Global Change* 6:1107417.
Giglio, L., L. Boschetti, D. P. Roy, M. L. Humber, and C. O. Justice. 2018. The collection 6 MODIS burned area mapping algorithm and product. *Remote Sensing of Environment* 217:72–85.
Ghorbanzadeh, O., T. Blaschke, K. Gholamnia, and J. Aryal. 2019. Forest fire susceptibility and risk mapping using social/infrastructural vulnerability and environmental variables. *Fire* 2 (3): 50.
Gong, H., J. Sardans, H. Huang, Z. Yan, Z. Wang, and J. Peñuelas. 2024. Global patterns and controlling factors of tree bark C: N: P stoichiometry in forest ecosystems consistent with biogeochemical niche hypothesis. *New Phytologist* 244 (4): 1303–1314.
Guo, S., B. Deng, C. Chen, J. Ke, J. Wang, S. Long, and K. Xu. 2024. Seeking in ride-on-demand service: A reinforcement learning model with dynamic price prediction. *IEEE Internet of Things Journal* 11 (18): 29890–29910.
Haider, K., M. F. Khokhar, F. Chishtie, W. RazzaqKhan, and K. R. Hakeem. 2017. Identification and future description of warming signatures over Pakistan with special emphasis on evolution of CO 2 levels and temperature during the first decade of the twenty-first century. *Environmental Science and Pollution Research* 24:7617–7629.
Hossen, S., Ali, Y., Chakma, S., Datta, S. and Hoque, A.R., 2024. Spatiotemporal analysis of land cover change, projection, and fragmentation: an application of Google earth engine and machine learning approach on Baraitali forest, Bangladesh. *Geology, Ecology, and Landscapes.* pp.1–14.
Hussain, K., Mehmood, K., Anees, S.A., Ding, Z., Muhammad, S., Badshah, T., Shahzad, F., Haidar, I., Wahab, A., Ali, J. and Ansari, M.J., 2024a. Assessing forest fragmentation due to land use changes from 1992 to 2023: A spatio-temporal analysis using remote sensing data. *Heliyon.* 10(14). https://doi.org/10.1016/j.heliyon.2024.e34710
Hussain, K., Mehmood, K., Yujun, S., Badshah, T., Anees, S.A., Shahzad, F., Nooruddin, Ali, J. and Bilal, M., 2024b. Analysing LULC transformations using remote sensing data: insights from a multilayer perceptron neural network approach. *Annals of GIS.* pp.1–27. https://doi.org/10.1080/19475683.2024.2343399
Ingwersen, E. W., W. T. Stam, B. J. V. Meijs, J. Roor, M. G. Besselink, B. Groot Koerkamp, I. H. J. T. de Hingh, H. C. van Santvoort, M. W. J. Stommel, and F. Daams. 2023. Machine learning versus logistic regression for the prediction of complications after pancreatoduodenectomy. *Surgery*. https://doi.org/10.1016/j.surg.2023.03.012.
Jalal, N., A. Mehmood, G. S. Choi, and I. Ashraf. 2022. A novel improved random forest for text classification using feature ranking and optimal number of trees. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2022.03.012.
Jallat, H., M. F. Khokhar, K. A. Kudus, M. Nazre, N. U. Saqib, U. Tahir, and W. R. Khan. 2021. Monitoring carbon stock and land-use change in 5000-year-old juniper forest stand of Ziarat, Balochistan, through a synergistic approach. *Forests* 12 (1): 51.
Jiao, Y., G. Zhu, S. Lu, L. Ye, D. Qiu, G. Meng, Qinqin Wang, Rui Li, Longhu Chen, Yuhao Wang, Dehong Si, and W. Li. 2024. The cooling effect of oasis reservoir-riparian forest systems in arid regions. *Water Resources Research* 60 (10) : e2024WR038301. https://doi.org/10.1029/2024WR038301.
Kalantar, B., N. Ueda, M. O. Idrees, S. Janizadeh, K. Ahmadi, and F. Shabani. 2020. Forest fire susceptibility prediction based on machine learning models with resampling algorithms on remote sensing data. *Remote Sensing* 12 (22): 3682.
Khan, W. R., F. Rasheed, S. Z. Zulkifli, M. R. B. M. Kasim, M. Zimmer, A. M. Pazi, N. A. Kamrudin, Z. Zafar, I. Faridah-Hanum, and M. Nazre. 2020. Phyto-extraction potential of *Rhizophora apiculata*: A case study in Matang mangrove forest reserve, Malaysia. *Tropical Conservation Science* 13 : 1940082920947344.
Khan, W. R., M. Nazre, S. Akram, S. A. Anees, K. Mehmood, F. H. Ibrahim, S. S. O. A. Edrus, A. Latiff, Z. A. Fitri, M. Yaseen, and P. Li. 2024. Assessing the productivity of the Matang mangrove forest reserve: Review of one of the best-managed mangrove forests. *Forests* 15 (5): 747.
Khasawneh, M. A., H. I. Al-Akhrass, S. R. Rabab'ah, and A. O. Al-sugaier. 2024. Prediction of California bearing ratio using soil index properties by regression and machine-learning techniques. *International Journal of Pavement Research and Technology* 17 (2): 306–324.
Kolanek, A., M. Szymanowski, and A. Raczyk. 2021. Human activity affects forest fires: The impact of anthropogenic factors on the density of forest fires in Poland. *Forests* 12 (6): 728.
Laha, A., N. Balasubramanian, and R. Sinha. 2021. Application of earth observation dataset and multi-criteria decision-making technique for forest fire risk assessment in Sikkim, India. *Current Science*. https://doi.org/10.18520/cs/v121/i8/1022-1031.
Lai, Z., J. Chao, Z. Zhao, M. Wen, H. Yang, W. Chai, Y. Yao, X. Zhao, Q. Chen, and J. Liu. 2023. Relationship between crustal deformation and thermal anomalies in the 2022 Ninglang Ms 5.5 earthquake in China: Clues from InSAR and RST. *Remote Sensing* 15 (5) : 1271.
Lamat, R., M. Kumar, A. Kundu, and D. Lal. 2021. Forest fire risk mapping using analytical hierarchy process (AHP) and earth observation datasets: A case study in the mountainous terrain of Northeast India. *SN Applied Sciences* 3 (4): 425.
Latha, S. 2022. Assessment of forest fire vulnerability zones in Nilgiris district, Tamilnadu, using remote sensing and GIS. *Bulletin of Environment, Pharmacology and Life Sciences* 11:67–87.
Li, R., G. Zhu, S. Lu, G. Meng, L. Chen, Y. Wang, E. Huang, Y. Jiao, and Q. Wang. 2025. Effects of cascade hydropower stations on hydrologic cycle in Xiying river basin, a runoff in Qilian mountain. *Journal of Hydrology* 646 : 132342.
Li, F., Y. Shan, M. Li, Y. Guo, and C. Liu. 2024. Insights for river restoration: The impacts of vegetation canopy length and canopy discontinuity on riverbed evolution. *Water Resources Research* 60 (7) : e2023WR036473.
Liu, J., Y. Zuo, N. Wang, F. Yuan, X. Zhu, L. Zhang, J. Zhang, Y. Sun, Z. Guo, Y. Guo, X. Song, C. Song, and X. Xu. 2021. Comparative analysis of two machine learning algorithms in predicting site-level net ecosystem exchange in major biomes. *Remote Sensing*. https://doi.org/10.3390/rs13122242.
Luo, M., S. A. Anees, Q. Huang, X. Qin, Z. Qin, J. Fan, G. Han, L. Zhang, and H. Z. M. Shafri. 2024. Improving forest above-ground biomass estimation

by integrating individual machine learning models. *Forests* 15 (6): 975. https://doi.org/10.3390/f15060975.
Luo, Q., C. Zhao, G. Luo, C. Li, C. Ran, S. Zhang, L. Xiong, J. Liao, C. Du, Z. Li, and Y. Xue. 2025. Rural depopulation has reshaped the plant diversity distribution pattern in China. *Resources, Conservation and Recycling* 215 : 108054.
Mallinis, G., M. Petrila, I. Mitsopoulos, A. Lorenţ, Ş Neagu, B. Apostol, V. Gancz, I. Popa, and J. G. Goldammer. 2019. Geospatial patterns and drivers of forest fire occurrence in Romania. *Applied Spatial Analysis and Policy*. https://doi.org/10.1007/s12061-018-9269-3.
Manzo-Delgado, L., S. Sánchez-Colón, and R. Álvarez. 2009. Assessment of seasonal forest fire risk using NOAA-AVHRR: A case study in central Mexico. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431160902852796.
McCandless, T. C., B. Kosovic, and W. Petzke. 2020. Enhancing wildfire spread modelling by building a gridded fuel moisture content product with machine learning. *Machine Learning: Science and Technology* 1 (3): 035010.
Mehmood, K., Anees, S.A., Rehman, A., Rehman, N.U., Muhammad, S., Shahzad, F., Liu, Q., Alharbi, S.A., Alfarraj, S., Ansari, M.J. and Khan, W.R. 2024c. Assessment of Climatic Influences on Net Primary Productivity along Elevation Gradients in Temperate Ecoregions. *Trees, Forests and People.* p.100657.
Mehmood, K., S. A. Anees, A. Rehman, A. Tariq, M. Zubair, Q. Liu, F. Rabbi, K. A. Khan, and M. Luo. 2024e. Exploring spatiotemporal dynamics of NDVI and climate-driven responses in ecosystems: Insights for sustainable management and climate resilience. *Ecological Informatics*. https://doi.org/10.1016/j.ecoinf.2024.102532.
Mehmood, K., S. A. Anees, A. Rehman, A. Tariq, Q. Liu, S. Muhammad, F. Rabbi, S. A. Pan, and W. A. Hatamleh. 2024d. Assessing forest cover changes and fragmentation in the Himalayan Temperate Region: Implications for forest conservation and management. *Journal of Forestry Research* 35 (1): 82. https://doi.org/10.1007/s11676-024-01734-6.
Mehmood, K., S. A. Anees, M. Luo, M. Akram, M. Zubair, K. A. Khan, and W. R. Khan. 2024a. Assessing Chilgoza Pine (*Pinus gerardiana*) forest fire severity: Remote sensing analysis, correlations, and predictive modeling for enhanced management strategies. *Trees, Forests, and People*. https://doi.org/10.1016/j.tfp.2024.100521.
Mehmood, K., S. A. Anees, S. Muhammad, F. Shahzad, Q. Liu, W. R. Khan, M. Shrahili, M. J. Ansari, and T. Dube. 2025a. Machine learning and spatio temporal analysis for assessing ecological impacts of the billion tree afforestation project. *Ecology and Evolution* 15 (2) : e70736.
Mehmood, K., S. A. Anees, S. Muhammad, G. Albasher, F. Shahzad, K. Hussain, Q. Liu, R. Ayub, and W. R. Khan. 2025b. Spatial and temporal vegetation dynamics from 2000 to 2023 in the Western Himalayan regions. *Stochastic Environmental Research and Risk Assessment*. https://doi.org/10.1007/s00477-025-02971-9.
Mehmood, K., S. A. Anees, S. Muhammad, K. Hussain, F. Shahzad, Q. Liu, M. J. Ansari, S. A. Alharbi, and W. R. Khan. 2024b. Analyzing vegetation health dynamics across seasons and regions through NDVI and climatic variables. *Scientific Reports* 14 (1) : 11775. https://doi.org/10.1038/s41598-024-62464-7.
Muhammad, S., A. Hamza, M. A. Kaleem Mehmood, and M. Tayyab. 2023. Analyzing the impact of forest harvesting ban in northern temperate forest. A case study of Anakar Valley, Kalam Swat Region, Khyber-Pakhtunkhwa, Pakistan. *Pure and Applied Biology (PAB)* 12 (2): 1434–1439.
Naskar, S., Rahaman, A., & Biswas, B. 2022. Forest Fire Susceptibility Mapping of West Sikkim District, India using MCDA techniques of AHP & TOPSIS model.
Michael, Y., D. Helman, O. Glickman, D. Gabay, S. Brenner, and I. M. Lensky. 2021. Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2020.142844.
Mohajane, M., R. Costache, F. Karimi, Q. B. Pham, A. Essahlaoui, H. Nguyen, G. Laneve, and F. Oudija. 2021. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecological Indicators* 129 : 107869.
Pandey, P., G. Huidobro, L. F. Lopes, A. Ganteaume, D. Ascoli, C. Colaco, G. Xanthopoulos, T. M. Giannaros, R. Gazzard, G. Boustras, and T. Steelman. 2023. A global outlook on increasing wildfire risk: Current policy situation and future pathways. *Trees, Forests, and People* 14 : 100431.
Ozenen Kavlak, M., S. N. Cabuk, and M. Cetin. 2021. Development of forest fire risk map using geographical information systems and remote sensing capabilities: Ören case. *Environmental Science and Pollution Research* 28 (25): 33265–33291.
Pan, S. A., S. A. Anees, X. Li, X. Yang, X. Duan, and Z. Li. 2023. Spatial and temporal patterns of non-structural carbohydrates in Faxon fir (*Abies fargesii* var. *faxoniana*), subalpine mountains of Southwest China. *Forests* 14 (7) : 1438. https://doi.org/10.3390/f14071438.
Phelps, N., and J. L. Beverly. 2022. Classification of forest fuels in selected fire-prone ecosystems of Alberta, Canada—implications for crown fire behaviour prediction and fuel management. *Annals of Forest Science* 79 (1): 40.
Pourtaghi, Z. S., H. R. Pourghasemi, and M. Rossi. 2015. Forest fire susceptibility mapping in the Minudasht forests, Golestan province, Iran. *Environmental Earth Sciences* 73 (4): 1515–1533.
Pramanick, N., B. Kundu, R. Acharyya, and A. Mukhopadhyay. 2023. Forest fire risk zone mapping in Mizoram using RS and GIS. *IOP Conference Series: Earth and Environmental Science* 1164 (1) : 012005.
Prasertsri, N., and P. Littidej. 2020. Spatial environmental modeling for wildfire progression accelerating extent analysis using geo-informatics. *Polish Journal of Environmental Studies*. https://doi.org/10.15244/pjoes/115175.
Rabby, Y. W., M. B. Hossain, and J. Abedin. 2022. Landslide susceptibility mapping in three Upazilas of Rangamati hill district Bangladesh: Application and comparison of GIS-based machine learning methods. *Geocarto International* 37 (12): 3371–3396.
Reyes-Bueno, F., and J. Loján-Córdova. 2022. Assessment of three machine learning techniques with open-access geographic data for forest fire susceptibility monitoring—Evidence from southern Ecuador. *Forests*. https://doi.org/10.3390/f13030474.
Ricotta, C., S. Bajocco, D. Guglietta, and M. Conedera. 2018. Assessing the influence of roads on fire ignition: Does land cover matter? *Fire* 1 (2) : 24. https://doi.org/10.3390/fire1020024.
Rodriguez-Galiano, V., M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas. 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews* 71:804–818.
Saidi, S., A. B. Younes, and B. Anselme. 2021. A GIS-remote sensing approach for forest fire risk assessment: Case of Bizerte region. *Tunisia. Applied Geomatics.* 13 (4): 587–603.
Shahzad, F., K. Mehmood, K. Hussain, I. Haidar, S. A. Anees, S. Muhammad, J. Ali, M. Adnan, Z. Wang, and Z. Feng. 2024. Comparing machine learning algorithms to predict vegetation fire detections in Pakistan. *Fire Ecology.* 20 (1): 1–20. https://doi.org/10.1186/s42408-024-00289-5.
Sharma, L. K., R. Gupta, and N. Fatima. 2022. Assessing the predictive efficacy of six machine learning algorithms for the susceptibility of Indian forests to fire. *International Journal of Wildland Fire* 31 (8): 735–758.
Shekar, P. R., A. Mathew, F. F. B. Hasher, K. Mehmood, and M. Zhran. 2025. Towards sustainable development: Ranking of soil erosion-prone areas using morphometric analysis and multi-criteria decision-making techniques. *Sustainability* 17 (5): 2124.
Shobairi, S. O. R., H. Lin, V. A. Usoltsev, A. A. Osmirko, I. S. Tsepordey, Z. Ye, and S. A. Anees. 2022. A comparative pattern for *Populus* spp. and *Betula* spp. stand biomass in Eurasian climate gradients. *Croatian Journal of Forest Engineering : Journal for Theory and Application of Forestry Engineering* 43 (2): 457–467. https://doi.org/10.5552/crojfe.2022.1340.
Singh, G., Kumar, P., & Mukherjee, S. 2023. Forest Fire Risk Mapping in Pauri Garhwal district of Uttarakhand.
Singha, C., K. C. Swain, A. Moghimi, F. Foroughnia, and S. K. Swain. 2024. Integrating geospatial, remote sensing, and machine learning for climate-induced forest fire susceptibility mapping in Similipal Tiger Reserve, India. *Forest Ecology and Management* 555 : 121729.
Su, H., W. Shen, J. Wang, A. Ali, and M. Li. 2020. Machine learning and geostatistical approaches for estimating aboveground biomass in Chinese subtropical forests. *Forest Ecosystems* 7:1–20.
Sun, W., S. Liang, G. Xu, H. Fang, and R. Dickinson. 2008. Mapping plant functional types from MODIS data using multisource evidential reasoning. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2007.07.022.

Tariq, A., H. Shu, S. Siddiqui, I. Munir, A. Sharifi, Q. Li, and L. Lu. 2022. Spatio-temporal analysis of forest fire events in the Margalla Hills, Islamabad, Pakistan using socio-economic and environmental variable data with machine learning methods. *Journal of Forestry Research* 33 (1): 183–194.
Tianqi, C., & Carlos, G. 2016. XGBoost: A Scalable Tree Boosting System Tianqi. The Journal of the Association of Physicians of India. 42(8):785-794.
Usoltsev, V. A., B. Chen, S. O. R. Shobairi, I. S. Tsepordey, V. P. Chasovskikh, and S. A. Anees. 2020. Patterns for *Populus* spp. stand biomass in gradients of winter temperature and precipitation of Eurasia. *Forests* 11 (9) : 906. https://doi.org/10.3390/f11090906.
Usoltsev, V. A., H. Lin, S. O. R. Shobairi, I. S. Tsepordey, Z. Ye, and S. A. Anees. 2022. The principle of space-for-time substitution in predicting *Betula* spp. biomass change related to climate shifts. *Applied Ecology and Environmental Research* 20 (4): 3683–3698. https://doi.org/10.15666/aeer/2004_36833698.
Wai, P., H. Su, and M. Li. 2022. Estimating aboveground biomass of two different forest types in Myanmar from sentinel-2 data with machine learning and geostatistical algorithms. *Remote Sensing* 14 (9): 2146.
Wang, R., S. Lu, and Q. Li. 2019. Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings. *Sustainable Cities and Society* 49 : 101623.
Xu, Y., D. Li, H. Ma, R. Lin, and F. Zhang. 2022. Modeling forest fire spread using machine learning-based cellular automata in a GIS environment. *Forests* 13 (12): 1974.
Yan, F., X. Wang, C. Huang, J. Zhang, F. Su, Y. Zhao, and V. Lyne. 2023. Sea reclamation in mainland China: Process, pattern, and management. *Land Use Policy* 127 : 106555.
Yathish, H., K. V. Athira, K. Preethi, U. Pruthviraj, and A. Shetty. 2019. A comparative analysis of forest fire risk zone mapping methods with expert knowledge. *Journal of the Indian Society of Remote Sensing*. https://doi.org/10.1007/s12524-019-01047-w.
Zafar, Z., M. S. Mehmood, A. Akbar, and M. A. Khan. 2023. Spatiotemporal dynamics analysis of surface water body and snow cover area to climate change in Gilgit Baltistan, Pakistan. *Physical Geography*. https://doi.org/10.1080/02723646.2023.2188633.
Zhang, F., B. Zhang, J. Luo, H. Liu, Q. Deng, L. Wang, and Z. Zuo. 2023a. Forest fire driving factors and fire risk zoning based on an optimal parameter logistic regression model: A case study of the Liangshan Yi Autonomous Prefecture, China. *Fire* 6 (9) : 336.
Zhang, X., M. Usman, A. U. R. Irshad, M. Rashid, and A. Khattak. 2024. Investigating spatial effects through machine learning and leveraging explainable AI for child malnutrition in Pakistan. *ISPRS International Journal of Geo-Information* 13 (9): 330.
Zhang, Y., Z. Gao, X. Wang, and Q. Liu. 2022. Predicting the pore-pressure and temperature of fire-loaded concrete by a hybrid neural network. *International Journal of Computational Methods* 19 (08): 2142011.
Zhang, Y., Z. Gao, X. Wang, and Q. Liu. 2023b. Image representations of numerical simulations for training neural networks. *Computer Modeling in Engineering & Sciences* 134 (2): 821–833.
Zhang, Y., X. Wang, C. Li, Y. Cai, Z. Yang, and Y. Yi. 2018. NDVI dynamics under changing meteorological factors in a shallow lake in future metropolitan, semiarid area in North China. *Scientific Reports*. https://doi.org/10.1038/s41598-018-33968-w.

Publisher's Note