

Article

Predicting Wildfire Susceptibility in Tanzanian Miombo Woodlands: A Random Forest-Based Spatio-Temporal Assessment in Iringa

John Rogath John ¹, Hui Huang ², Haifeng Gao ³, Xiaoying Han ⁴, Faris Jamal Mohamedi ¹, Abbas Khurram ¹, Xiangxuan Zeng ^{5,*} and Zhan Shu ^{1,*}

¹ College of Forestry, Northeast Forestry University, Harbin 150040, China; johnrogath@nefu.edu.cn (J.R.J.); mohamedifar@nefu.edu.cn (F.J.M.); khurramabbass64@nefu.edu.cn (A.K.)

² School of Water Conservancy and Civil Engineering, Heilongjiang Agricultural Engineering Vocational College, Harbin 150025, China; swkjhh@163.com

³ Satellite Application Center for Ecology and Environment, Ministry of Ecology and Environment, Beijing 100006, China; gaohf03@hotmail.com

⁴ Institute of Atmospheric Environment and Climate Change, Heilongjiang Academy of Environmental Sciences, Harbin 150000, China; 18646250280@163.com

⁵ Department of International Economics and Trade, Cheongju University, Cheongju 360-764, Republic of Korea

* Correspondence: zengxiangxuan@kakao.com (X.Z.); shuzhan@nefu.edu.cn (Z.S.); Tel.: +86-451-82192030 (Z.S.)

Abstract

Wildfires threaten natural ecosystems and human livelihoods in the Tanzanian Miombo woodlands. This study presents the first locally calibrated, high-resolution wildfire susceptibility map for the Iringa region, developed using a robust machine learning framework. Multi-decadal remote sensing data (MODIS fire occurrences, 2001–2024) were integrated with climatic, topographic, vegetation, and anthropogenic variables to train four classifiers: Random Forest, XGBoost, support vector machine with RBF kernel, and Logistic Regression. A balanced dataset of 9096 fire points and an equal number of randomly sampled non-fire points was used. The data were split into 70% for training and 30% for testing. Model performance was evaluated using accuracy, area under the ROC curve (AUC), accuracy, precision, and F1-score. Random Forest achieved the highest overall performance (AUC = 0.845, accuracy = 0.759, precision = 0.789 and F1 = 0.771), followed by XGBoost (AUC = 0.828, accuracy = 0.736, precision = 0.700 and F1 = 0.757), SVM (AUC = 0.755, accuracy = 0.679, precision = 0.648 and F1 = 0.709), and Logistic Regression (AUC = 0.740, accuracy = 0.661, precision = 0.631 and F1 = 0.696). Feature importance analysis identified altitude as the most influential variable, followed by wind speed, distance to road, and NDVI. Kernel Density Estimation revealed spatially distinct fire clusters concentrated in central and southern hotspots. Temporal analysis showed that 94% of fires occur during the dry season (June–November), peaking sharply in October. These findings provide an evidence-based framework for fire prevention and sustainable management of Iringa's Miombo woodlands.

Keywords: Miombo woodlands; wildfire prediction; wildfire susceptibility; Random Forest; machine learning

Academic Editors: Fuquan Zhang and Yiqing Xu

Received: 19 April 2026

Revised: 19 June 2026

Accepted: 6 July 2026

Published: 9 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

As important components of tropical savanna environments, southern Africa's Miombo woodlands are crucial for regulating regional climate, sequestering carbon, and supporting rural livelihoods [1]. Dominated by deciduous tree genera such as *Brachystegia*, *Julbernardia*, *Pterocarpus angolensis*, and *Isoberlinia*, they form the most extensive woodland ecosystem in Africa, covering approximately 2.7 million km² across southern, central, and eastern Africa [2,3]. These tree species exhibit strong drought resistance through deep root systems and water-conservative physiological traits [4].

In Tanzania, the Miombo biome represents the country's largest vegetation type and is highly fire-prone due to a prolonged dry season of seven to eight months, which generates abundant grassy fuels that promote frequent wildfires [5,6]. In the Iringa region, fire acts as a major ecological force shaping woodland structure and function, influencing biodiversity, carbon storage, and human safety. Growing climatic variability and human activity raise concerns about future changes in fire frequency and severity [7–9]. To move from reactive fire suppression to proactive risk-based landscape management, it is essential to understand the spatial and temporal dynamics of wildfire and to develop accurate, locally calibrated prediction tools [10].

Wildfires emerge from complex, nonlinear interactions among meteorological conditions, fuel characteristics, topography, and human ignition sources [11]. Climate conditions regulate fuel properties and thus fire potential [12], while topography critically governs fire spread by modulating wind fields and fire–atmosphere interactions [13]. At the same time, anthropogenic land-use practices can strongly shape fire seasonality and may even override climatic controls in some regions [14]. Modern fire activity is therefore driven by a combination of environmental forcing and human influence, making wildfire prediction inherently complex [15–17]. A reliable wildfire susceptibility model must account for these multiple interacting drivers [18].

Given this complexity, machine learning has become increasingly valuable in environmental modeling because it can capture nonlinear relationships in ecological data [19]. Forest fire susceptibility mapping is commonly treated as a binary classification problem, and ensemble algorithms, particularly Random Forest, have proven especially effective by improving robustness and reducing predictive variance [20–22]. However, recent studies have demonstrated the utility of comparing multiple algorithms, including XGBoost, Support Vector Machine with RBF kernel, and Logistic Regression, to ensure robust predictions across diverse environmental settings [23,24]. Random Forest constructs a multitude of decision trees during training and outputs the mode of the classes for classification tasks [25]. Its strong predictive performance, stability across different regions, and ability to rank variable importance make it well-suited for the complex, multidimensional nature of environmental datasets [26,27]. These data-driven models overcome major limitations of traditional approaches by objectively learning from historical patterns, enabling superior predictive accuracy and the capacity to process high-dimensional datasets [10].

Despite global progress, critical research gaps persist in the Miombo woodlands, one of Africa's most fire-prone biomes [28]. First, although global and regional fire models exist, their application to data-scarce regions like the Miombo woodlands remains limited; effective use requires local calibration with ground-truthed data [29]. Second, regional studies often prioritize maximizing model accuracy with multi-factor datasets, thereby obscuring how core meteorological variables independently drive observed fire patterns. Third, high-resolution, long-term analyses of East African wildfire dynamics are notably scarce, particularly those that connect predictive model outputs to actionable risk maps for fire management authorities [7].

Addressing these gaps requires a deliberate modeling strategy. Although socio-economic factors such as land-use practices, ignition behavior, and fire management are widely recognized as important determinants of wildfire activity [8,9], including them directly in a predictive model can make it difficult to isolate the specific influence of biophysical conditions. Accordingly, the methodological approach of this study is not to construct a fully comprehensive model that captures all dimensions of human activity, but rather to establish a robust biophysical baseline that quantifies the independent contribution of meteorological conditions alongside key topographic, vegetative, and static anthropogenic variables [30]. This baseline serves as a clean starting point, providing a foundational framework upon which future coupled analyses can build by integrating dynamic socio-economic and land-use factors to more fully represent the human dimensions of fire regimes in Miombo woodlands [31].

This study focuses on the Miombo woodlands of Tanzania's Iringa region. Multiple machine learning classifiers were developed and validated using local wildfire records and recent remote sensing products to construct a baseline biophysical framework for wildfire susceptibility prediction. The study has three objectives: (1) to develop and evaluate multiple machine learning models (Random Forest, XGBoost, Support Vector Machine, and Logistic Regression) for wildfire susceptibility mapping in the Iringa region, (2) to analyze the spatio-temporal distribution of wildfire occurrence from 2001 to 2024, and (3) to identify and quantify the main biophysical drivers including climatic, topographic, and vegetation factors associated with wildfire occurrence. By integrating these components, the study provides an interpretable tool to support wildfire prevention, resource allocation, and sustainable management of a vulnerable but ecologically important socio-ecological system.

2. Materials and Methods

2.1. Study Area Description

The Iringa region is located in the Southern Highlands of Tanzania, with geographical coordinates extending from 6°55' to 9°00' South and 33°45' to 36°55' East; see Figure 1. The region is bordered by the Singida and Dodoma Regions to the north, Morogoro to the east, Mbeya to the west, and Njombe to the south. Encompassing an area of about 35,743 km², the region's terrain is mainly land (33,038.8 km²), with water bodies covering the remaining 2704.2 km² [32]. Nearly three-quarters of the land is arable; the rest comprises grazing lands, natural forests, game reserves, plantation forests, farmlands and wetlands. The region is characterized by a subtropical highland climate with marked wet and dry seasons, with annual average rainfall typically ranging between 500 mm and 1600 mm, and mean annual temperatures generally ranging from 11 °C to 28 °C. The area is dominated by fire-adapted Miombo woodlands such as *Brachystegia boehmii* and *Julbernardia globiflora*, alongside montane forests, grasslands, and increasingly extensive agricultural areas [33]. Fire occurrence in this region follows a distinct seasonal pattern driven by the local climate, human activities, vegetation cover and topographic factors [6].

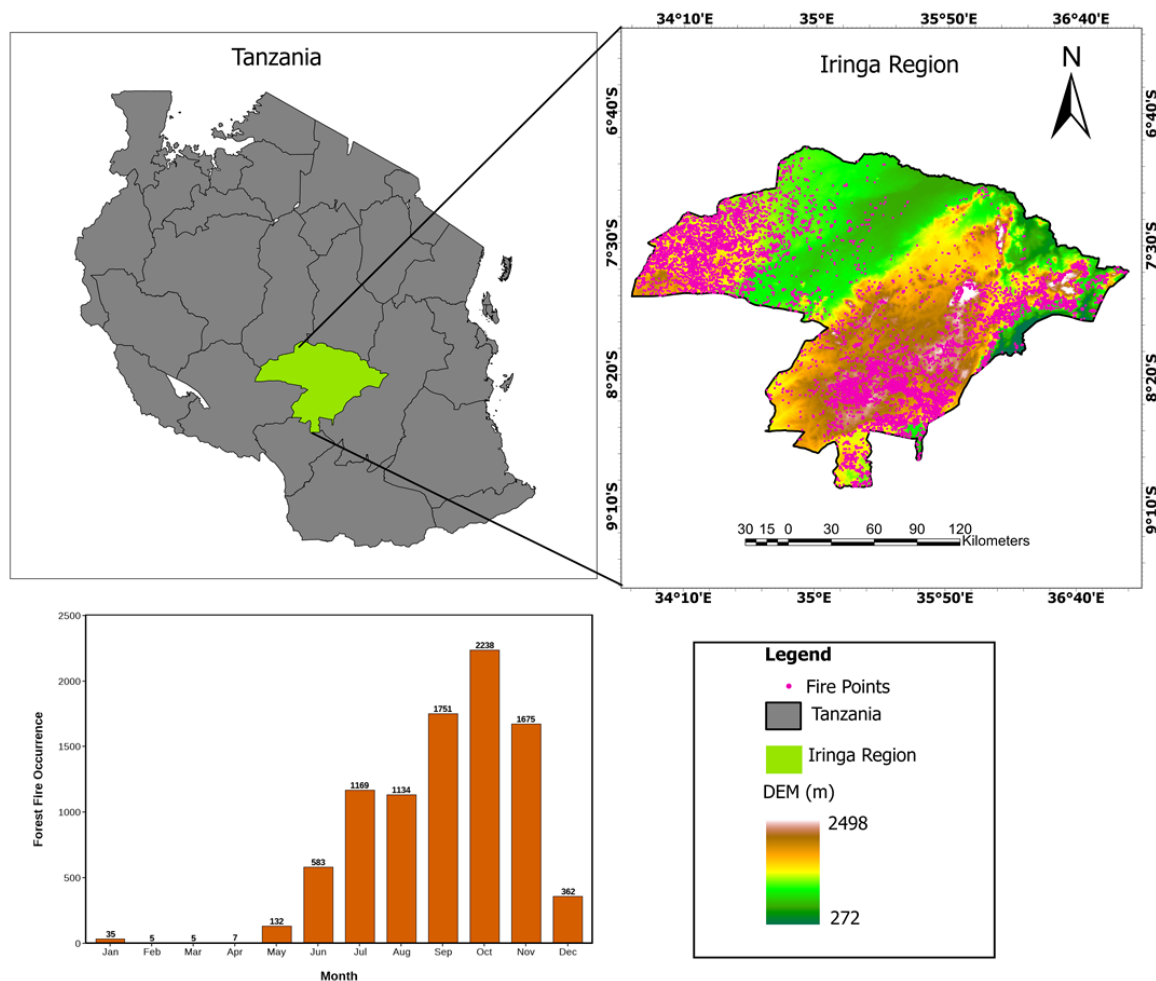


Figure 1. This map depicts the Iringa region of Tanzania, highlighting historical fire occurrences from 2001 to 2024. Fire locations are shown as pink points overlaid on topography, with elevations spanning 272 to 2498 m. An embedded histogram details the monthly distribution of fires.

2.2. Data Resources

2.2.1. Forest Fire Data

Our study used the MODIS active fire data product (MCD14ML) provided by NASA Fire Information for Resource Management (FIRMS) spanning 2001 to 2024 with 1 km resolution [34]. Fire occurrence data were processed in Google Earth Engine and filtered to retain only detections with a confidence level of at least 95%. The MODIS Land Cover Type product (MCD12Q1) was subsequently used to exclude non-target land covers, ensuring that only fire events occurring within the specified vegetation types were retained [35]. This strict confidence threshold prioritizes reliability and helps minimize false positives in the final analysis [36,37]. Large-scale comparisons confirm that MODIS systematically misses smaller or low-intensity fires despite high-confidence filtering [38]. After data filtering, a total of 9096 valid fire-point instances were identified. To construct a balanced dataset, a 1 km spatial buffer was created around each fire point, and random non-fire points were sampled exclusively from outside these buffers, ensuring that non-fire locations were not positioned immediately adjacent to actual fire events (where similar environmental conditions might prevail). Each selected non-fire point was assigned a date randomly drawn from the temporal distribution of the 9096 fire points [18,39,40]. This temporal sampling strategy guarantees that fire and non-fire observations share comparable seasonal and interannual climatic variability [41]. Together, these spatial and temporal procedures produce a representative set of locations and times where fires did not

occur, despite similar environmental opportunities for ignition. The final analytical dataset comprised 18,192 complete observations, evenly split between 9096 fire events and 9096 non-fire events.

2.2.2. Predictor Variables

In this study we selected 12 predictor variables, which were categorized into climatic (daily temperature, windspeed, humidity, and precipitation), vegetative (NDVI, vegetation type), topographic (altitude, slope and aspect), and anthropogenic (distance to road, distance to river, population density), as shown in Figure 2 and Table 1.

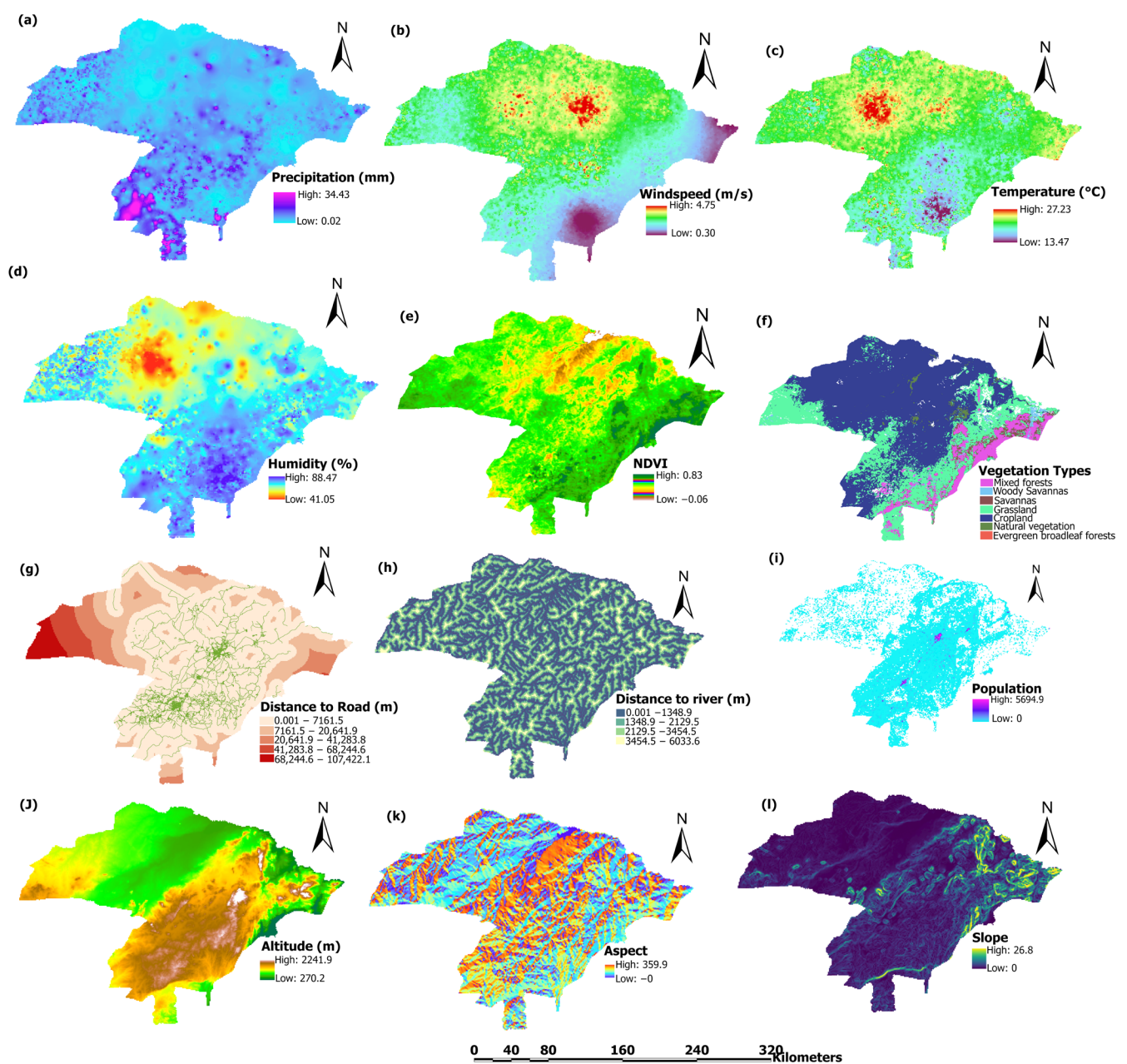


Figure 2. Factors influencing forest fires in the Iringa region, including climatic factors (a–d), vegetation factors (e,f), anthropogenic factors (g–i) and topographic factors (j–l); the figure was created using ArcGIS Pro 4.5.

(1) The climatic variables, including daily temperature, wind speed, precipitation, and humidity, were obtained from the NASA Prediction of Worldwide Energy Resources (POWER) project as daily raster datasets at 2 m spatial resolution and were extracted to match the exact date of each fire occurrence, ensuring that the meteorological conditions reflected the day of the event [42]. (2) For the topographic variables, a Digital Elevation Model (DEM), from which altitude, slope, and aspect were derived, was obtained from the Shuttle Radar Topography Mission (SRTM) at a spatial resolution of 30 m [43]. (3) Anthropogenic variables, including distance to road, distance to rivers, and population density, were obtained from open street map road network [44], hydrosheds hydrology dataset [45], and Tanzania national bureau statistics, respectively. (4) Vegetation characteristics were assessed using MODIS-derived NDVI and land cover classification, both at 500 m resolution. All variables were harmonized to a consistent spatial framework using ArcGIS Pro (Version 4.5). For continuous datasets for climate variables, DEM-derived indices, NDVI, population density, and distance-based metrics we applied Inverse Distance Weighting (IDW) interpolation to generate continuous raster surfaces. IDW was selected for its computational efficiency and robustness with moderately dense sampling points, and it remains widely adopted for environmental variable interpolation [46]. To reconcile the disparate native resolutions of these layers, we resampled all continuous rasters to a uniform 1 km pixel size using bilinear interpolation. This resolution is a practical compromise between preserving ecologically meaningful spatial detail and maintaining manageable data volumes for regional-scale analysis [47,48].

For categorical vegetation type, we did not apply IDW; instead, we resampled it using the nearest-neighbor method to preserve the original class assignments without creating artificial transitional values [49]. Finally, multicollinearity among all predictors was assessed using variance inflation factor (VIF) and tolerance thresholds, confirming that no problematic redundancy existed prior to model training [50].

Table 1. Description of the forest fires influencing factors.

Influencing Factor	Predictor Variable	Unit	Source
Meteorology	Windspeed	m/s	NASA POWER-Data Access Viewer
	Humidity	%	
	Temperature	°C	
	Precipitation	mm	
Topography	Altitude	M	STRM
	Slope	Degree	
	Aspect	Degree	
Human activities	Population Density	Persons/km ²	National Bureau Statistics (Tanzania)
	Distance to Road	M	OpenStreetMap (OSM) road network
	Distance to River	M	https://www.hydrosheds.org/ (accessed on 11 March 2026)
Vegetation	Normalized difference vegetation	Ratio	MODIS
	Index (NDVI)	(scale −1 to 1)	
	Vegetation Type	Class	

2.3. Machine Learning Model Application

A comparison of four algorithms (Random Forest (RF), XGBoost, Support Vector Machine (SVM) with RBF kernel, and Logistic Regression (LR)) was performed to ensure robust wildfire prediction [51,52]. An identical preprocessing and validation pipeline was applied to all models. The dataset was divided by stratified random sampling into a training set (70%) and a held-out testing set (30%), thereby preserving class proportions [53]. To prevent data leakage, the two validation strategies were applied sequentially; the test

set was fully withheld and was not used in any hyperparameter tuning or cross-validation step. Specifically, a 5-fold stratified cross-validation (CV) was performed exclusively on the 70% training set for hyperparameter tuning [54]. For the Random Forest, three parameters were tuned via this CV: the number of candidate variables per split (*mtry*), the minimum samples required to split a node (*min.node.size*), and the maximum tree depth (*max.depth*). The optimal configuration was selected by maximizing the CV F1-score [55]. After tuning, the final RF model was retrained on the full training set with 500 trees and *class_weight* = *balanced_subsample*; an out-of-bag score was reported. XGBoost was trained with 300 trees (learning rate 0.05), SVM with an RBF kernel and class balancing, and LR with balanced class weights. All models received the same scaled features (RobustScaler). When class imbalance exceeded 20%, the Synthetic Minority Over-sampling Technique (SMOTE) was applied only to the training set [55]. The optimal probability threshold for each model was determined by maximizing the F1-score on validation predictions. Final performance was then evaluated once on the never-seen 30% test set using AUC-ROC, F1-score, accuracy, recall, specificity and precision.

2.3.1. Performance Evaluation Metrics

Model performance was evaluated using the following complementary metrics, all calculated from the confusion matrix [18]:

Let *TP* = True Positives, *TN* = True Negatives, *FP* = False Positives, and *FN* = False Negatives.

Accuracy measured overall prediction correctness (Equation (1)):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Precision quantified the proportion of correctly predicted fire events (Equation (2)):

$$Precision = TP / (TP + FP) \quad (2)$$

Recall (Sensitivity) measured the proportion of actual fire events correctly identified (Equation (3)):

$$Recall = TP / (TP + FN) \quad (3)$$

F1-score provided the harmonic mean of precision and recall (Equation (4)):

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

AUC represented the probability that the model ranks a randomly chosen positive instance higher than a randomly chosen negative instance (Equation (5)):

$$AUC = \int_0^1 TPR(FPR^{-1}(x))dx \quad (5)$$

where TPR = True Positive Rate (Recall) and FPR = False Positive Rate (1—Specificity).

2.3.2. Variable Importance Analysis

To enhance the interpretability of the model, we assessed variable importance to quantify the relative contribution of each predictor to the final forest fire prediction output [26]. The predictor variable importance was derived from the internal model structure, based on the contribution of each predictor based on node splitting and Gini impurity reduction. To evaluate the independent contribution of each environmental and topographical predictor to the forest fire model, permutation importance was calculated following the approach implemented in the R package *vip* (R Version 4.5). This model-agnostic technique avoids the known bias of Gini-based importance toward correlated or

multi-category predictors, making it a robust choice for comparing the influence of different variable types in wildfire susceptibility mapping [56,57].

2.4. Correlation and Multicollinearity Assessment of Predictor Variables

To examine the relationships among the predictor variables (climatic, topographic, vegetative, and anthropogenic) and to detect potential multicollinearity prior to machine learning modeling, a two-step correlation and diagnostic approach was implemented [51,52]. First, Pearson correlation coefficients (r) were computed to evaluate linear associations among all continuous predictors, while Spearman rank correlation (ρ) was used to assess monotonic relationships, particularly for variables with non-normal distributions or ordinal characteristics (vegetation type) [58]. Second, multicollinearity was formally diagnosed using the variance inflation factor (VIF) and tolerance ($\text{TOL} = 1/\text{VIF}$). For each predictor, VIF was calculated through an ordinary least squares regression of that variable against all remaining predictors. Following established thresholds, VIF values below 10 and TOL values above 0.1 were considered indicative of acceptable multicollinearity levels [52,59]. All analyses were performed in RStudio using the `cor.test`, `cor`, and `vif` functions (from the `car` package). The correlation matrices were visualized using `corrplot` with hierarchical clustering, and VIF results were tabulated for transparent reporting. No variable transformations were applied for correlation analyses, as the subsequent Random Forest algorithm is non-parametric and robust to non-normality [26].

2.5. Spatial Pattern Analysis Using Kernel Density Estimation

To identify areas of persistent fire activity, we converted point-based wildfire records into a continuous hotspot surface using Kernel Density Estimation (KDE) in ArcGIS Pro. This widely adopted method is effective for detecting spatial clusters and has a strong precedent in both fire science and ecological research [60]. The primary data source was the Fire Points feature class, which contains all georeferenced fire incident data sourced from MODIS active fire [61]. To ensure accurate area measurements for calculating density, we projected all data into the WGS 1984 UTM zone 36S coordinate reference system.

The Kernel Density Estimation method was applied to convert a discrete fire point into a continuous density surface, computing fire event magnitude per unit area within a defined neighborhood [62]. A search radius of 5000 m was selected to optimally delineate the local fire clusters. The analysis was performed at an output cell size of 500 m, providing detail for landscape-level assessment of the study area. The resulting fire density raster, expressed as events per square kilometer, was classified using the Natural Breaks (Jenks) method. This approach identifies inherent groupings in the data by minimizing variance within classes and maximizing variance between them. We divided the density values into five intuitive categories: Extremely Low, Low, Medium, High, and Extremely High. A sequential color ramp progressing from cool greens to intense reds was applied to visually distinguish areas of low to high fire density, allowing for clear and immediate interpretation.

3. Results

3.1. Model Performance Evaluation and Comparative Assessment

3.1.1. Model Performance Comparison

Among the models evaluated, Random Forest (RF) yielded the highest overall predictive performance for forest fire susceptibility mapping (Table 2). As shown in Figure 3, RF achieved the highest AUC-ROC (0.845), followed by XGBoost (0.828), SVM (RBF) (0.755), and Logistic Regression (0.740), indicating strong discriminative ability of the ensemble tree-based methods. Random Forest also attained the highest accuracy (0.759), pre-

cision (0.789), specificity (0.812), and F1-score (0.771). The Out-of-Bag (OOB) score of 0.775 closely matched the test accuracy of 0.759, confirming model consistency and generalizability to unseen data. Random Forest correctly detected 70.6% of actual fire events (recall) while maintaining a low false-positive rate, reflected by its high specificity of 81.2%. Among the remaining models, XGBoost exhibited the highest recall (0.825) but lower specificity (0.647). SVM (RBF) and Logistic Regression showed moderate performance across most metrics. Overall, Random Forest consistently outperformed the other three classifiers across the majority of evaluation metrics, establishing it as the most suitable model for practical fire susceptibility assessment.

To further assess the temporal robustness of the Random Forest model and address the potential for optimistic performance inflation arising from the random train–test partitioning of spatially and temporally autocorrelated wildfire data, a temporally independent holdout evaluation was conducted. The dataset was partitioned chronologically, with observations from 2001 to 2019 used for model training and observations from 2020 to 2024 reserved as a strictly independent test set, ensuring that no future fire event information influenced model training. The temporally trained model achieved an OOB score of 0.773, an AUC-ROC of 0.806, an F1-score of 0.733, and a Balanced Accuracy of 0.689 on the temporal test set (Table 2). The difference between the random-split AUC-ROC (0.845) and the temporal holdout AUC-ROC (0.806) was 0.039, which falls below the commonly adopted threshold of 0.05 used to flag substantial autocorrelation-induced bias. This small and acceptable margin indicates that the model’s discriminative performance is temporally stable and that the random-split estimates are not substantially inflated by temporal autocorrelation between training and test observations. The consistency between the two evaluation strategies therefore lends additional credibility to the reported performance metrics and supports the broader generalisability of the Random Forest model for wildfire susceptibility assessment in the Iringa region.

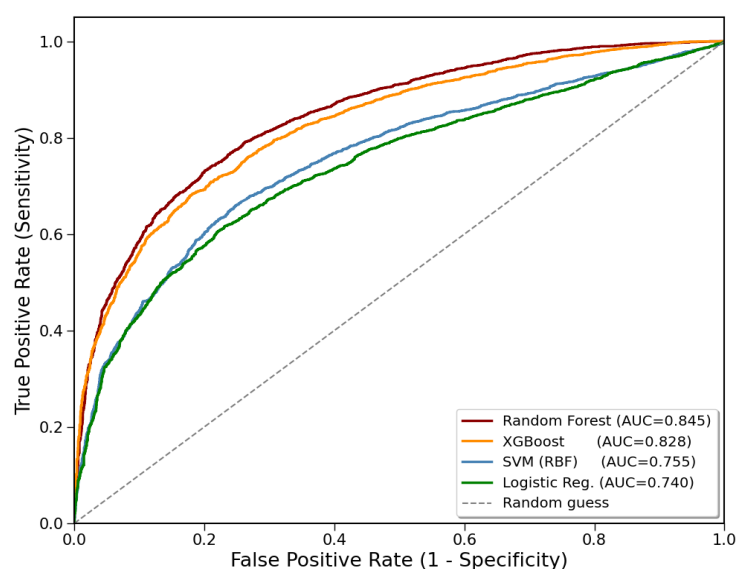


Figure 3. Receiver Operating Characteristic (ROC) curve for the Random Forest, XGBoost, SVM and Logistic Regression, illustrating the model’s discriminatory capability between fire and non-fire conditions.

Table 2. Performance comparison of four classifiers (Random Forest, XGBoost, SVM with RBF kernel, and Logistic Regression) on the test set.

Metric	RF	XGBoost	SVM (RBF)	Logistic Regression	RF Temporal Holdout
Accuracy	0.759	0.736	0.679	0.661	-
Precision	0.789	0.700	0.648	0.631	-
Recall	0.706	0.825	0.782	0.776	-

Specificity	0.812	0.647	0.576	0.546	-
F1-score	0.771	0.757	0.709	0.696	0.733
AUC-ROC	0.845	0.828	0.755	0.740	0.806
Balanced Accuracy	-	-	-	-	0.689
OOB Score	0.775	-	-	-	0.773

3.1.2. Wildfire Susceptibility Map

The wildfire susceptibility map for the Iringa region from 2001 to 2024 was generated by the Random Forest model; see Figure 4. The map reveals a spatially heterogeneous wildfire susceptibility classified using the natural break method, with very high to high susceptibility concentrated in southern, western, and peripheral areas, identifying priority hotspots for targeted fire prevention and management. Very-high-susceptibility zones dominate the landscape (34.99%) and exhibit the highest mean probability (0.898) shown in Table 3, indicating strong fire potential. High and moderate classes together account for 19.11% of the area, with intermediate probabilities. In contrast, low- and very-low-susceptibility zones comprise 45.89% of points and are characterized by consistently low fire probabilities, reflecting comparatively stable and less fire-prone conditions.

Table 3. Summary of wildfire susceptibility classification results.

Susceptibility Class	Number of Points	Percentage (%)	Average Probability	Probability Range
Very High	6366	35.0	0.90	0.80–1.00
High	2413	13.3	0.72	0.60–0.80
Moderate	1065	5.9	0.51	0.40–0.60
Low	2427	13.3	0.27	0.20–0.40
Very Low	5921	32.6	0.09	0.00–0.20

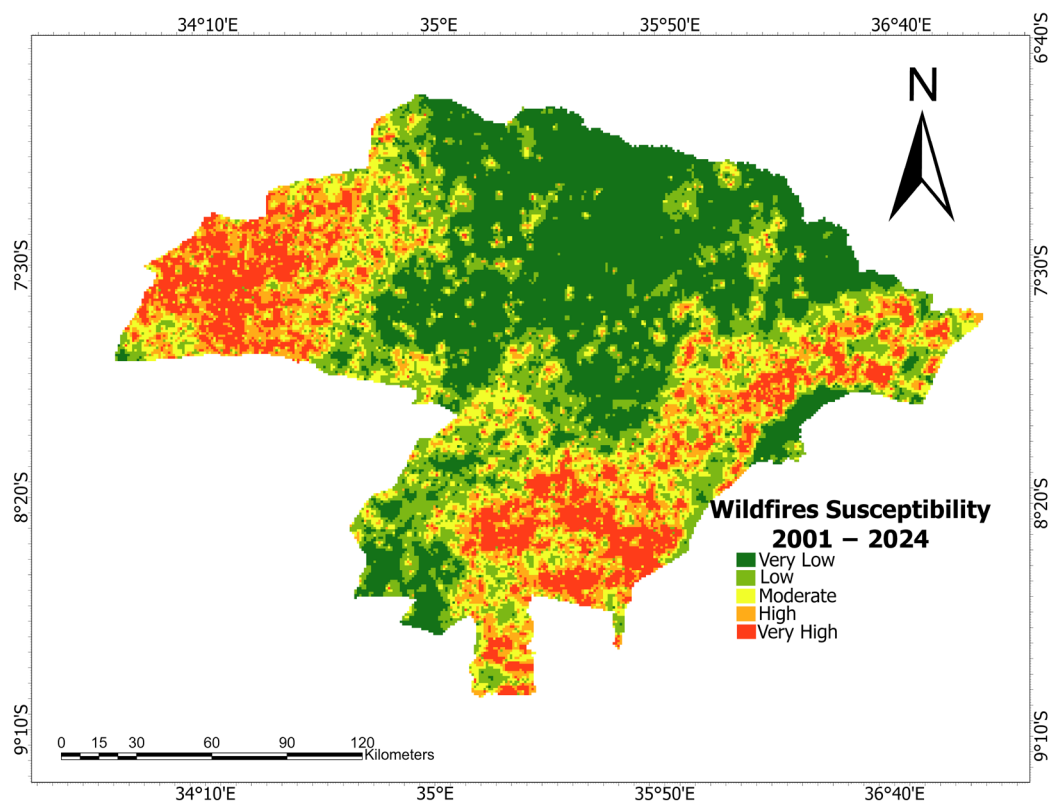


Figure 4. Wildfire susceptibility map of the Iringa region classified using the natural breaks (Jenks) method in ArcGIS Pro 4.5.

3.1.3. Variable Importance Analysis

Feature importance analysis based on mean decrease in Gini impurity identified a clear hierarchy among the predictor variables in their contribution to wildfire occurrence classification (Figure 5). Altitude emerged as the dominant predictor (Gini importance = 0.173), accounting for the largest proportion of node purity improvement across the 500 decision trees, likely reflecting its integrative influence on temperature lapse rates, fuel moisture content, and fire weather conditions that co-vary systematically with elevation. Wind speed ranked second (0.131), consistent with its well-established mechanistic role in fire spread through the drying of surface fuels and the transport of firebrands. Distance to road (0.094) and NDVI (0.085) occupied the third and fourth positions respectively, underscoring the dual significance of anthropogenic ignition proximity and vegetation greenness, a proxy for fuel load and moisture status in determining fire susceptibility. Humidity (0.076), precipitation (0.074), slope (0.073), and temperature (0.073) exhibited closely comparable importance scores, suggesting that these meteorological and topographic variables contribute relatively equally and may act synergistically rather than independently in conditioning fire risk. Distance to river (0.062) and population density (0.061) contributed modestly, indicating secondary roles of hydrological boundaries and human presence within the study landscape. Aspect (0.061) and vegetation type (0.034) ranked lowest, with the limited contribution of vegetation type potentially reflecting the dominance of continuous variables in discriminating fire from non-fire conditions, or homogeneity in vegetation classes across the study area.

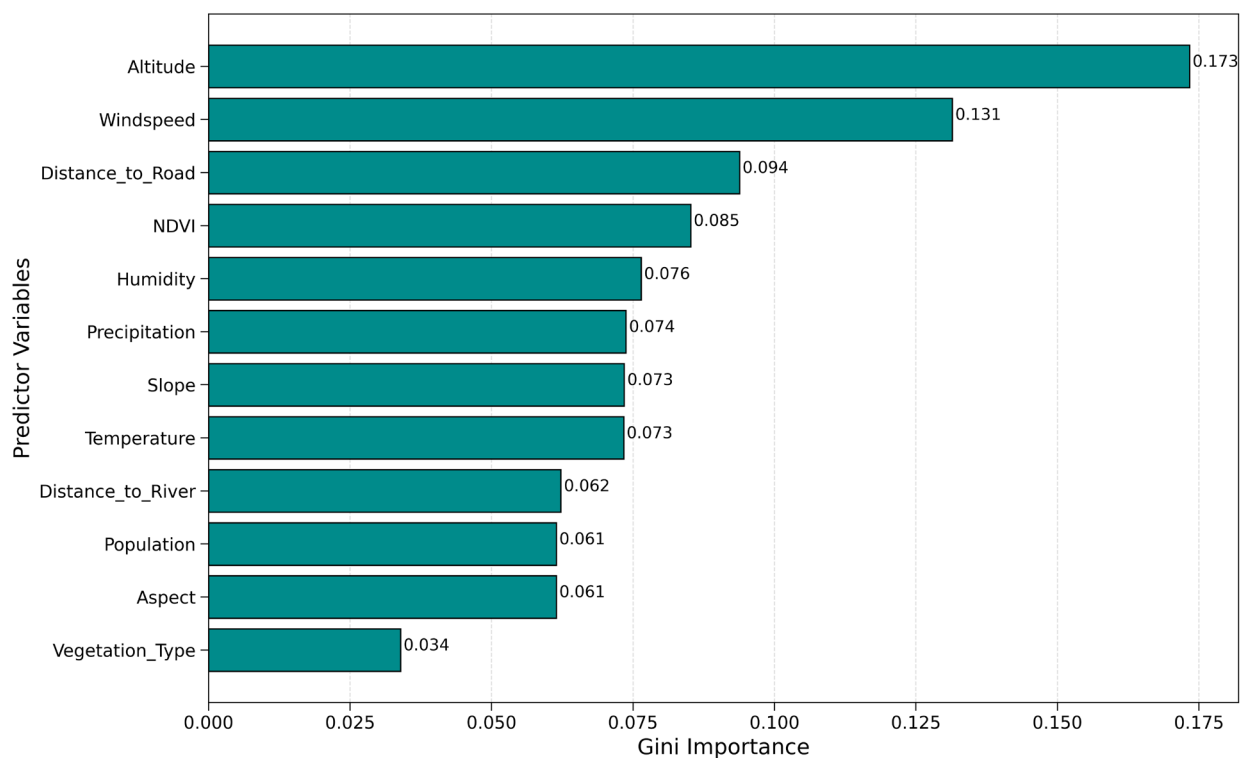


Figure 5. Feature importance ranking showing the relative contributions of predictor variables to fire prediction.

It should be noted, however, that Gini impurity-based importance is known to exhibit bias toward continuous, high-cardinality predictors a property that may partially account for the elevated ranking of altitude [63]. To verify the robustness of this importance hierarchy, permutation importance was additionally computed on the held-out test set (Figure S1). Re-assuringly, both methods converged on a similar top-ranking structure, with altitude, wind

speed, distance to road, NDVI, and humidity consistently occupying the top five positions, confirming that their dominant influence on wildfire occurrence reflects genuine predictive signal rather than a methodological artifact of the Gini measure.

3.2. Correlation Analysis

3.2.1. Pearson Correlation Analysis

Pearson correlations (Figure 6 highlight how climatic, topographic and vegetation variables interact to create fire-prone conditions in Miombo woodlands. The strongest relationship is a negative link between wind speed and humidity ($r = -0.64$), meaning that windy days are typically dry. This combination directly accelerates fuel drying and makes fire spread easier. Wind speed also rises with temperature ($r = 0.43$), so the hottest periods are also the windiest. Together, these two patterns explain why late dry-season conditions are so hazardous: heat, dryness and wind reinforce each other.

Humidity's negative correlations with both wind speed and temperature ($r = -0.64$ and -0.57) are ecologically critical: when relative humidity falls, both heat and wind tend to increase, creating a synergistic drop in fuel moisture. Precipitation correlates only moderately with humidity ($r = 0.43$) and very weakly with other variables. Hence, rainfall amount alone is a poor fire predictor; the timing of rain and the length of dry spells matter far more.

Vegetation (NDVI) is strongly shaped by wind and topography. NDVI drops as wind speed rises ($r = -0.52$), meaning that wind-exposed areas support sparser, drier vegetation, a self-reinforcing feedback that increases fire danger. NDVI is also positively associated with humidity ($r = 0.31$) and altitude ($r = 0.37$). Consequently, higher, moister sites have greener, denser fuels that are less flammable, while lower, windier valleys carry more continuous, dry fuels that promote fire.

Altitude itself shows a clear protective effect: it correlates positively with humidity and NDVI, but negatively with temperature ($r = -0.40$) and wind speed ($r = -0.31$). Higher elevations are cooler, moister and less windy, which reduces fire susceptibility. Slope and aspect show weak inter-correlations, meaning they contribute largely independent information; slope's moderate association with NDVI ($r = 0.29$) suggests steeper terrain may retain more woody biomass.

Anthropogenic variables (population, distances to road and river) have very low linear correlations with environmental factors ($|r| < 0.15$). This is methodologically important: human influences on fire regimes (e.g., farming near roads) operate largely independently of natural gradients, so their effects can be isolated in the predictive models without confounding.

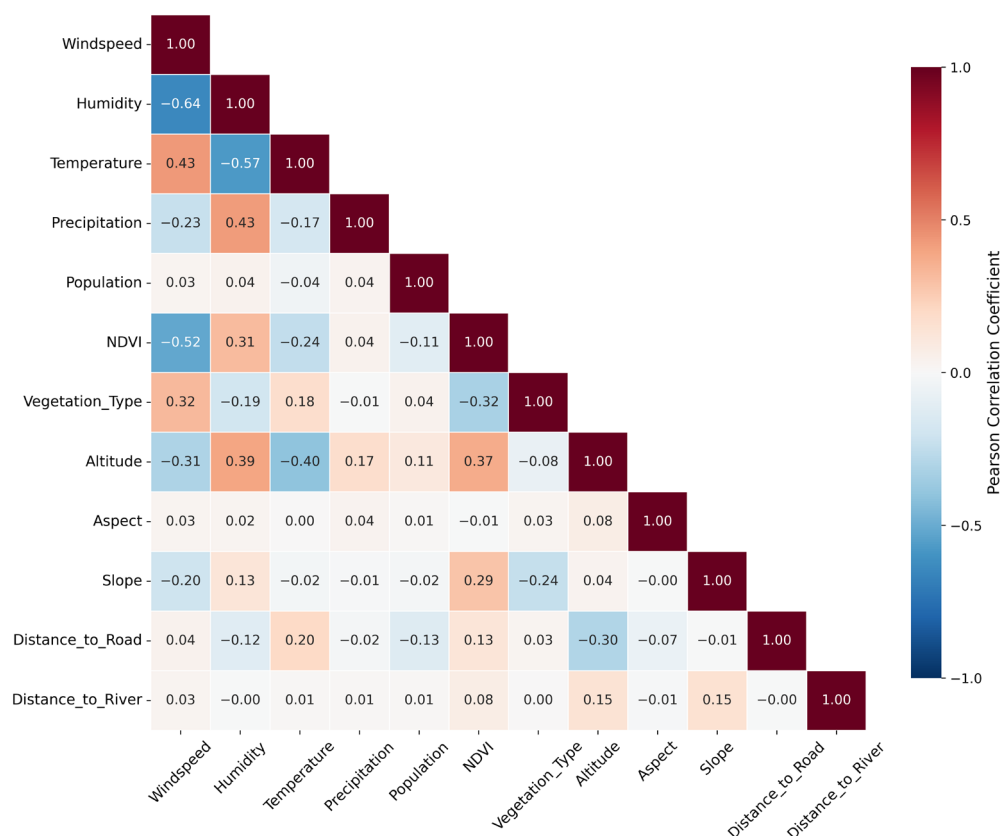


Figure 6. Pearson correlation matrix showing the relationships among the predictor variables.

3.2.2. Spearman Correlation Analysis

Spearman rank correlations (Figure 7) confirm the monotonic patterns and add information for nonlinear or ordinal variables, especially vegetation type. The strong inverse relationship between wind speed and humidity ($\rho = -0.65$) again underlines that fire risk rises when dry air and strong winds coincide.

Vegetation type (ordinal) is strongly negatively correlated with NDVI ($\rho = -0.64$). This reflects that more open, grass-dominated classes have lower NDVI and are typically more flammable than closed woodlands. Wind speed also correlates negatively with NDVI ($\rho = -0.57$), reinforcing that windier areas support sparser vegetation, a positive feedback for fire spread.

The negative association between population density and distance to road ($\rho = -0.53$) is directly relevant for fire occurrence: densely populated areas lie closer to roads, which are common ignition sources (agricultural burning, discarded cigarettes, accidental sparks). Fire risk is therefore expected to be higher near road networks in populated zones.

Moderate positive correlations include vegetation type with wind speed ($\rho = 0.54$): higher ordinal values (more open, dry vegetation classes) are associated with windier conditions. Altitude correlates positively with population density ($\rho = 0.43$), indicating that in the Iringa region, human settlements tend to occupy higher elevations—possibly for historical or climatic reasons. Yet altitude is also negatively correlated with temperature ($\rho = -0.44$). This trade-off matters for fire modeling: although higher elevations are cooler and less fire-prone, they host more people, so human-caused ignitions may partially offset the climatic advantage.

Humidity's positive correlation with precipitation ($\rho = 0.52$) and negative correlation with temperature ($\rho = -0.60$) reaffirm the climatic triangle: warm and dry conditions increase fire risk. Aspect shows negligible correlations with all other predictors ($| \rho | < 0.09$), meaning it provides completely independent information about solar exposure and fuel

drying—a valuable predictor that does not duplicate other variables. Importantly, no pairwise correlation exceeds $|\rho| = 0.70$, confirming that multicollinearity will not distort variable importance or model predictions in the subsequent machine learning analyses.

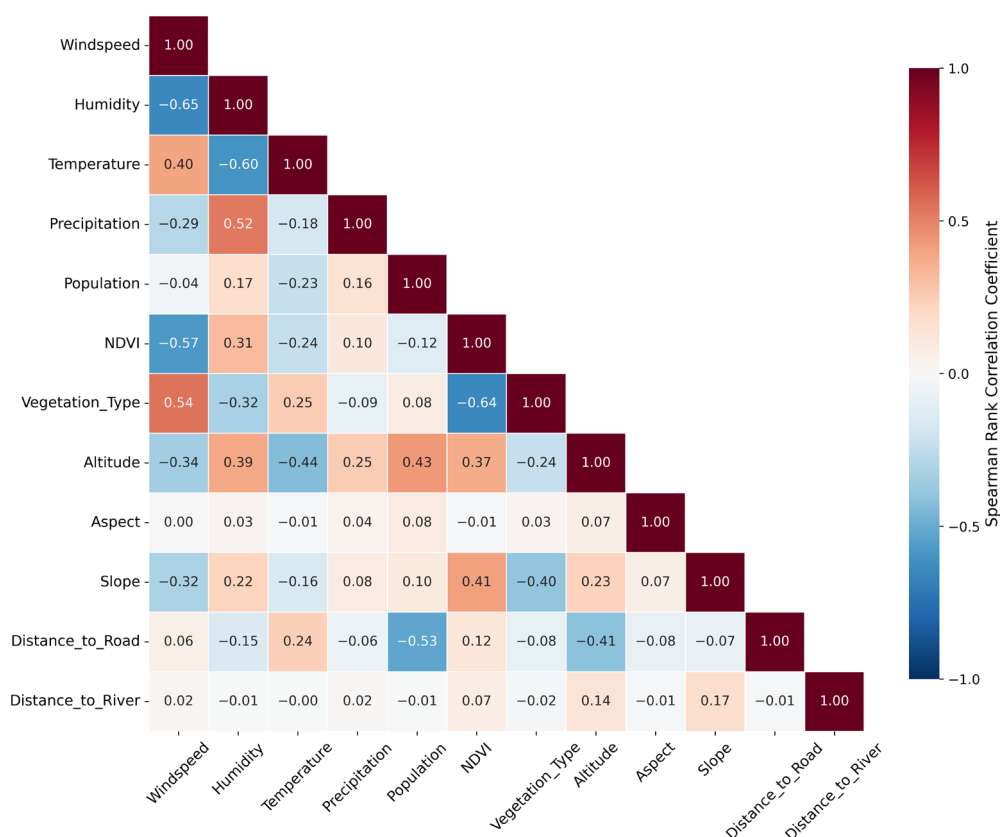


Figure 7. Spearman rank correlation matrix showing the relationships among the predictor variables.

3.3. Multicollinearity Analysis

To formally assess multicollinearity among the predictors, variance inflation factors (VIFs) and tolerance (TOL) values were computed for all 12 variables (Table 4). The results showed that VIF values ranged from 1.02 (Aspect) to 2.51 (Humidity), with corresponding TOL values ranging from 0.40 to 0.98. All variables comfortably satisfied the recommended thresholds ($\text{TOL} > 0.1$ and $\text{VIF} < 10$), confirming the absence of problematic multicollinearity. Consequently, all 12 predictors were retained for model training and validation, ensuring that the subsequent Random Forest analysis and other machine learning models would not be compromised by redundant information and that variable importance measures would reliably reflect independent contributions.

Table 4. Multicollinearity diagnosis of the predictor variables (features).

No.	Feature	VIF	TOL
1	Windspeed	2.36	0.42
2	Humidity	2.51	0.40
3	Temperature	1.65	0.61
4	Precipitation	1.27	0.79
5	Population	1.05	0.96
6	NDVI	1.90	0.53
7	Altitude	1.66	0.60
8	Aspect	1.02	0.98
9	Slope	1.22	0.82

10	Distance_to_Road	1.31	0.77
11	Distance_to_River	1.06	0.94
12	Vegetation_Type	1.95	0.51

3.4. Spatial Patterns of Wildfire Density

Our application of Kernel Density Estimation (KDE) delineates the pronounced spatial clustering of wildfires in Tanzania's Iringa region between 2001 and 2024, as shown in Figure 8. The resulting density surface, classified for interpretation, reveals distinct core hotspots alongside zones of minimal activity.

Concentrated areas of extremely high and high fire density form well-defined hotspots, primarily in the region's central and southeastern sectors. These clusters represent landscapes with the most frequent ignitions, indicating persistently high fire activity over the 24-year study period. Conversely, huge areas in the west and north are characterized by low to extremely low density, marking them as relatively fire-sparse.

The medium-density class typically creates a transitional buffer, often surrounding or connecting the high-density clusters and illustrating a gradient in fire frequency. This non-random, highly concentrated spatial pattern confirms that a disproportionate share of ignitions recur in specific locales. Such a distribution strongly suggests the influence of essential regional drivers, whether ecological, anthropogenic, or topographic, that predispose these areas to burning. Ultimately, the statistically significant hotspots identified here provide crucial spatial targets for investigating underlying meteorological causes and for prioritizing future fire management efforts.

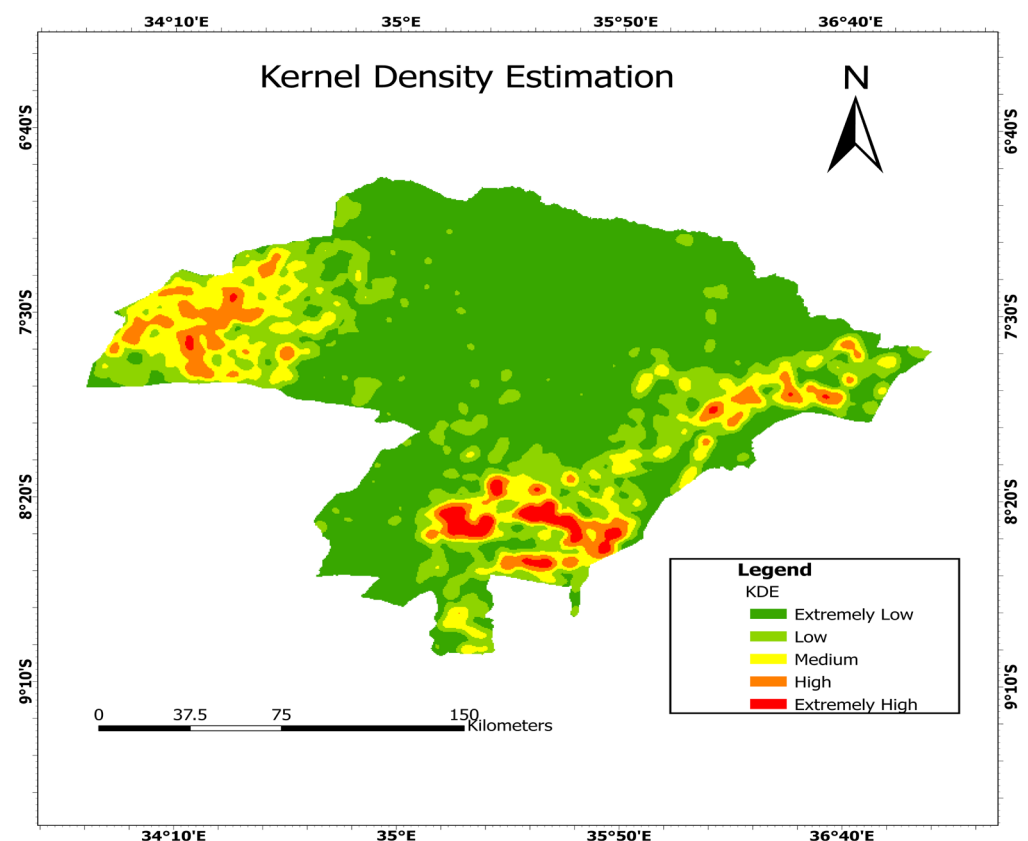


Figure 8. Spatial density of wildfire occurrences. Kernel density estimates identify geographic hotspots (red/orange) and low-frequency areas (green).

3.5. Temporal Distribution of Forest Fires

Between 2001 and 2024, a total of 9096 wildfires were recorded in the Iringa region, making an annual average of 379 fires, though the numbers varied significantly from year to year. The analysis pinpointed three peak years: 2021 saw the highest number with 712 fires, followed by 2010 (644 fires), and 2003 (627 fires), as illustrated in Figure 9.

Fire activity revealed a strong seasonal concentration, with 96.4% of all ignitions clustered between June and November. This establishes a clear dry-season fire regime for the region. October was consistently the peak month, accounting for 2238 fires (a mean of 93.25 per year), followed by September (1751 fires) and November (1675 fires). These details are summarized in Table 5.

Table 5. Summary statistics of forest fire occurrences.

Metric	Result
Total fires (2001–2024)	9096
Mean annual fires	379
Standard deviation	150.9
Peak fire year	2021
Fires in peak year	712
Peak fire month	October
Fires in peak month	2238
Dry season fires (Jun–Nov)	8550
% in dry season	94.00%

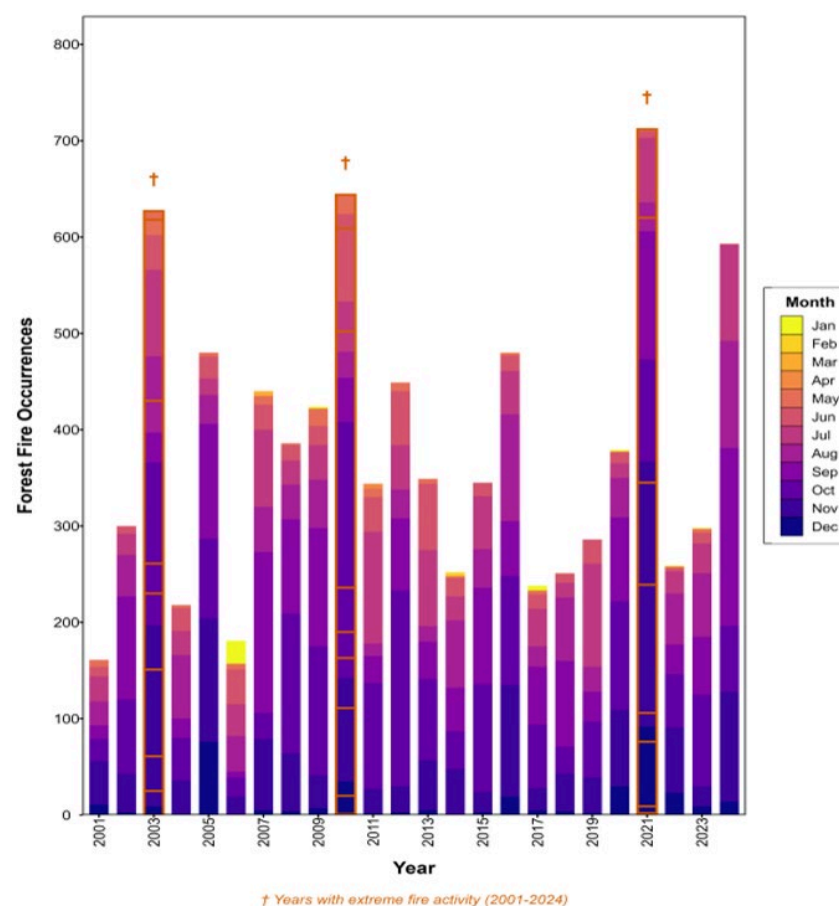


Figure 9. Showing the interannual and seasonal variability of forest fire occurrences.

4. Discussion

Wildfires are a growing global threat, making accurate prediction tools essential. To fill a gap in Tanzanian research, this study assesses wildfire risk in Miombo woodlands using several machine learning models: Random Forest, XGBoost, SVM, and Logistic Regression. In earlier work by [64], the Fire Weather Index was predicted at Sao Hill Forest Plantation using a BiGRU model (test RMSE = 0.065), while [65] applied a ConvLSTM to predict binary fire occurrence on Mount Kilimanjaro (AUROC = 0.9785, accuracy = 98.08%). In contrast, the Random Forest model employed in this study attained an accuracy of 0.759 and an AUC of 0.845. Although [65] achieved higher predictive performance, this study remains valuable because it addresses a larger, more heterogeneous Miombo woodland landscape (the Iringa region) with a comprehensive set of environmental and anthropogenic predictors, providing a more generalizable and operationally relevant fire susceptibility assessment for an area that has been historically understudied. Furthermore, the use of Random Forest (RF) helps resist overfitting, a known weakness of simpler algorithms such as decision trees [25]. The reliability of RF across different ecological settings was confirmed by its alignment with other global and African savanna applications [18,21].

The most influential factors controlling fire susceptibility in this understudied region were identified. Altitude was ranked as the top predictor, consistent with [66], who identified topography as a key driver of fire severity in the West Usambara Mountains. The influence of altitude on local microclimate, moisture availability, and vegetation structure is well-documented in Miombo woodlands [4,33]. Wind speed was ranked second, reflecting its well-established role in driving wildfire propagation and reducing moisture in ground fuels [12–14]. Distance to roads was ranked third, supporting the findings of [67] that transportation networks are the primary driver of ecological fires in Tanzania's Miombo woodlands. This relationship is attributed to increased human access facilitated by road infrastructure, enabling ignition-related activities such as agricultural land preparation (34.1%), hunting (28.5%), livestock grazing (9.3%), and charcoal production (0.7%) [68].

A practical contribution to a data-scarce region has been made. By combining 24 years of fire observations with spatially interpolated predictor variables, the lack of localized, high-resolution fire risk maps for African savannas noted by [29] was addressed. The integration of both anthropogenic and vegetation variables was found to be essential, as human-caused ignition and seasonal land use practices fundamentally drive fire regimes in Miombo woodlands [40,69]. An AUC of 0.845 is therefore interpreted as a realistically high performance for an operational model in a human-shaped, data-scarce landscape, where complex nonlinear interactions create a noisier predictive signal than in climate-dominated systems [10].

A fire susceptibility map was produced, identifying very-high- to high-risk areas predominantly in the southern, western, and peripheral regions. The accuracy of this map was validated by the close alignment of historical fire locations with these predictions, with significantly greater fire density observed within the very-high-risk class. The spatial distribution of fire susceptibility was shown to be driven by climate, anthropogenic pressures, and topography [70]. A pronounced dry season peak was confirmed, with 94–96% of fires occurring between June and November, primarily peaking in October. This pattern parallels southern African Miombo woodland fire regimes [6,9]. From a practical standpoint, September–October was highlighted as a critical window for targeted monitoring and early warning efforts.

Kernel Density Estimation was used to identify concentrated fire hotspots in central and southeastern Iringa. These persistent clusters were found to align with broader Miombo woodland patterns, where topography, human activity, and microclimate synergistically shape high-risk zones [60,62]. Hence, these zones are proposed as priority areas for

targeted prevention, including focused patrols, community-based fire management, and strategic fuel break placement [10,15].

In summary, wildfire risk management for Tanzania's Miombo woodlands has been advanced by locally calibrating Random Forest models with long-term MODIS data. The identification of persistent spatio-temporal hotspots provides a strategic basis for proactive resource allocation, shifting focus from reactive firefighting to predictive mitigation.

5. Limitation of the Study

Despite the robust findings presented, several limitations must be acknowledged when interpreting the results.

Firstly, the fire occurrence data used in this study were derived from the MODIS product at a 1 km spatial resolution. Although this resolution is suitable for regional-scale analysis, it was recognized that small-scale fires or accurate fire perimeters might not be detected by this product, potentially leading to an underestimation of fire frequency in areas where fine-scale burning practices dominate.

Secondly, certain predictor variables employed in the model, such as distance to roads, were treated as static. Consequently, the temporal dynamics of human activity intensity, including seasonal variations in land use, infrastructure development, or shifting patterns of access, could not be captured by this approach.

Thirdly, natural ignition sources, particularly lightning strikes, were not incorporated into the modeling framework due to the absence of high-quality, long-term lightning data for the study region. As a result, it is possible that some natural ignitions were inadvertently attributed to other correlated variables by the model.

Fourthly, the model was calibrated using historical data from 2001 to 2024, and stationarity in the relationships between drivers and fire occurrence was assumed. Therefore, its predictive performance under future climate change scenarios, characterized by altered precipitation regimes, increased temperature extremes, and shifts in vegetation composition, remains unknown and warrants further investigation.

Fifthly, the Inverse Distance Weighting (IDW) method was applied to spatially interpolate all predictor variables into continuous surfaces. However, the spatial variability of meteorological variables (e.g., temperature, precipitation, wind speed, and humidity) exhibits anisotropy and topographic dependence, characteristics that cannot be adequately captured by IDW. For these meteorological variables, more appropriate methods such as co-kriging or terrain-based interpolation would have been preferable. Hence, the implications of this methodological choice should be considered when evaluating the spatial accuracy of the climatic predictor layers, and future studies are encouraged to adopt co-kriging or terrain-based approaches for such variables in the Iringa region.

Sixthly, model interpretability and the absence of advanced explainability methods represent another limitation. Feature importance in this study was assessed using two complementary approaches, Gini impurity reduction and permutation importance, both of which produced a convergent top-five predictor ranking, thereby providing a robust and internally validated basis for global feature interpretation. However, both methods are inherently global measures that summarize average feature contributions across all predictions and do not reveal the directionality of individual feature effects, the nature of nonlinear relationships, or interaction effects between predictors. SHAP (SHapley Additive exPlanations), grounded in cooperative game theory, would have provided additive, locally consistent decomposition of individual predictions, enabling both global summary visualizations and observation-level force plots that reveal how each predictor pushes a specific prediction above or below the baseline. Similarly, LIME would have enabled locally faithful, interpretable explanations for individual pixel-level predictions, which would be particularly valuable for communicating fire risk to land managers and policy stakeholders. Future studies should incorporate Tree-

SHAP analysis to examine the direction, magnitude, and interaction structure of feature effects, particularly in operational deployment contexts where model transparency and stakeholder trust are essential requirements.

6. Conclusions

This work establishes that a Random Forest model, trained and tested on local data, can effectively map wildfire susceptibility across the Miombo woodlands of Iringa. The analysis points to altitude, wind speed, and road proximity as the principal controls on fire susceptibility not as statistical artifacts, but as factors with clear ecological and human dimensions. Altitude modulates local climate and fuel moisture; wind speed accelerates fuel drying and fire spread; and roads reflect the imprint of human activity on ignition patterns. Together, these variables capture the integrated natural and anthropogenic drivers that shape fire regimes in this landscape. The spatial patterns are instructive. Fire activity is not evenly distributed but concentrates in persistent hotspots across central and southeastern Iringa. Temporally, the dry season months from June through November account for the overwhelming majority of fires, with October standing out as the peak month. These findings translate directly into practical guidance: managers can now target seasonal patrols and community outreach to well-defined geographic areas where fire risk is consistently high.

This study provides the first empirically grounded, high-resolution wildfire susceptibility framework for the Iringa region, filling a notable gap in fire research for Tanzanian Miombo woodlands. The framework is designed to be replicable and updatable as new observational data become available. However, the operational applicability of the model should be interpreted with appropriate caution. While the model performs well against historical data, its predictive strength in real-world management contexts would benefit from additional validation against independent spatio-temporal datasets, particularly as climate conditions and land use continue to evolve. For the present, the susceptibility map offers a practical, evidence-based starting point for prioritizing fire prevention efforts and supporting sustainable forest management in this fire-prone socio-ecological system. Future work would do well to extend the temporal validation window and explore the integration of socioeconomic dynamics to further strengthen the framework's utility for long-term planning.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/fire9070289/s1>, Figure S1: Permutation importance score.

Author Contributions: Conceptualization, Z.S. and J.R.J.; methodology, J.R.J.; software, J.R.J.; validation, F.J.M., Z.S. and A.K.; formal analysis, J.R.J.; data curation, Z.S. and J.R.J.; writing—original draft preparation, J.R.J.; writing—review and editing, J.R.J. and F.J.M.; visualization, J.R.J. and F.J.M.; supervision, Z.S., X.H., H.G., X.Z. and H.H.; project administration, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (grant nos. 2024YFF1306200, 2024YFF1306203).

Data Availability Statement: The data available upon request from the corresponding author.

Acknowledgments: The authors would like to express their sincere gratitude to the College of Forestry, Northeast Forestry University, Harbin, China, for their financial support and the provision of necessary facilities. Finally, we would like to acknowledge the valuable and constructive comments provided by the anonymous reviewers, which have greatly improved the quality of this manuscript.

Conflicts of Interest: The author declares no conflicts of interest.

References

- Ribeiro, N.S.; Katerere, Y.; Chirwa, P.W.; Grundy, I.M. Introduction. In *Miombo Woodlands in a Changing Environment: Securing the Resilience and Sustainability of People and Woodlands*; Springer International Publishing: Cham, Switzerland, 2020; pp. 1–8.
- Backéus, I.; Pettersson, B.; Strömquist, L.; Ruffo, C. Tree Communities and Structural Dynamics in Miombo (*Brachystegia-Julbernardia*) Woodland, Tanzania. *For. Ecol. Manag.* **2006**, *230*, 171–178. <https://doi.org/10.1016/j.foreco.2006.04.033>.
- Campbell, B. *The Miombo in Transition: Woodlands and Welfare in Africa*; Center for International Forestry Research (CIFOR): Bogor, Indonesia, 1996.
- Vinya, R.; Malhi, Y.; Brown, N.D.; Fisher, J.B.; Brodribb, T.; Aragão, L.E.O.C. Seasonal Changes in Plant–Water Relations Influence Patterns of Leaf Display in Miombo Woodlands: Evidence of Water Conservative Strategies. *Tree Physiol.* **2019**, *39*, 104–112. <https://doi.org/10.1093/treephys/tpy062>.
- Russell-Smith, J.; Yates, C.; Vernooij, R.; Eames, T.; Lucas, D.; Mbindo, K.; Banda, S.; Mukoma, K.; Kaluka, A.; Liseli, A.; et al. Framework for a Savanna Burning Emissions Abatement Methodology Applicable to Fire-Prone Miombo Woodlands in Southern Africa. *Int. J. Wildland Fire* **2024**, *33*, WF23193. <https://doi.org/10.1071/WF23193>.
- Tarimo, B.; Dick, Ø.B.; Gobakken, T.; Totland, Ø. Spatial Distribution of Temporal Dynamics in Anthropogenic Fires in Miombo Savanna Woodlands of Tanzania. *Carbon Balance Manag.* **2015**, *10*, 18. <https://doi.org/10.1186/s13021-015-0029-2>.
- Meneses, O.M.; Ribeiro, N.S.; Shirvani, Z.; Andrew, S.M. Spatio-Temporal Analysis of Wildfire Regimes in Miombo of the LevasFlor Forest Concession, Central Mozambique. *Fire* **2024**, *7*, 264. <https://doi.org/10.3390/fire7080264>.
- Wood, E. *Causes and Consequences of Fire in the Miombo Woodlands of South-Eastern Tanzania*; Ph.D. Thesis, The University of Edinburgh, Edinburgh, UK, 2022.
- Villaverde Canosa, I.; Ford, J.; Paavola, J.; Burnasheva, D. Community Risk and Resilience to Wildfires: Rethinking the Complex Human–Climate–Fire Relationship in High-Latitude Regions. *Sustainability* **2024**, *16*, 957. <https://doi.org/10.3390/su16030957>.
- Mishra, M.; Guria, R.; Baraj, B.; Nanda, A.P.; Santos, C.A.G.; da Silva, R.M.; Laksono, F.A.T. Spatial Analysis and Machine Learning Prediction of Forest Fire Susceptibility: A Comprehensive Approach for Effective Management and Mitigation. *Sci. Total Environ.* **2024**, *926*, 171713. <https://doi.org/10.1016/j.scitotenv.2024.171713>.
- Patil, R.; Pawar, J.; Shah, K.; Shetty, D.; Ajith, A.; Jadhav, S. Machine Learning Based Forest Fire Prediction: A Comparative Approach. *Int. Res. J. Multidiscip. Technovation* **2024**, *6*, 32–39. <https://doi.org/10.54392/irjmt2413>.
- Jolly, W.M.; Freeborn, P.H.; Bradshaw, L.S.; Wallace, J.; Brittain, S. Modernizing the US National Fire Danger Rating System (Version 4): Simplified Fuel Models and Improved Live and Dead Fuel Moisture Calculations. *Environ. Model. Softw.* **2024**, *181*, 106181. <https://doi.org/10.1016/j.envsoft.2024.106181>.
- Linn, R.; Winterkamp, J.; Edminster, C.; Colman, J.J.; Smith, W.S. Coupled Influences of Topography and Wind on Wildland Fire Behaviour. *Int. J. Wildland Fire* **2007**, *16*, 183. <https://doi.org/10.1071/WF06078>.
- Le Page, Y.; Oom, D.; Silva, J.M.N.; Jönsson, P.; Pereira, J.M.C. Seasonality of Vegetation Fires as Modified by Human Action: Observing the Deviation from Eco-climatic Fire Regimes. *Glob. Ecol. Biogeogr.* **2010**, *19*, 575–588. <https://doi.org/10.1111/j.1466-8238.2010.00525.x>.
- Bowman, D.M.J.S.; Kolden, C.A.; Abatzoglou, J.T.; Johnston, F.H.; van der Werf, G.R.; Flannigan, M. Vegetation Fires in the Anthropocene. *Nat. Rev. Earth Environ.* **2020**, *1*, 500–515. <https://doi.org/10.1038/s43017-020-0085-3>.
- Kelly, L.T.; Fletcher, M.-S.; Oliveras Menor, I.; Pellegrini, A.F.A.; Plumanns-Pouton, E.S.; Pons, P.; Williamson, G.J.; Bowman, D.M.J.S. Understanding Fire Regimes for a Better Anthropocene. *Annu. Rev. Environ. Resour.* **2023**, *48*, 207–235. <https://doi.org/10.1146/annurev-environ-120220-055357>.
- Vigna, I.; Besana, A.; Comino, E.; Pezzoli, A. Application of the Socio-Ecological System Framework to Forest Fire Risk Management: A Systematic Literature Review. *Sustainability* **2021**, *13*, 2121. <https://doi.org/10.3390/su13042121>.
- Li, Y.; Li, G.; Wang, K.; Wang, Z.; Chen, Y. Forest Fire Risk Prediction Based on Stacking Ensemble Learning for Yunnan Province of China. *Fire* **2023**, *7*, 13. <https://doi.org/10.3390/fire7010013>.
- Hu, L.; Hochschild, V.; Neidhardt, H.; Schultz, M.; Khosravani, P.; Shokati, H. BIPE: A Bi-Layer Predictive Ensemble Framework for Forest Fire Susceptibility Mapping in Germany. *Remote Sens.* **2024**, *17*, 7. <https://doi.org/10.3390/rs17010007>.
- Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY. *Ecology* **2007**, *88*, 2783–2792. <https://doi.org/10.1890/07-0539.1>.
- Nguyen Van, L.; Lee, G. Optimizing Stacked Ensemble Machine Learning Models for Accurate Wildfire Severity Mapping. *Remote Sens.* **2025**, *17*, 854. <https://doi.org/10.3390/rs17050854>.
- Zha, M.; Wang, Z.; Ji, J.; Zhu, J. The Distributed Strategy for Asynchronous Observations in Data-Driven Wildland Fire Spread Prediction. *Int. J. Wildland Fire* **2024**, *33*. <https://doi.org/10.1071/WF23165>.

23. Moghim, S.; Mehrabi, M. Wildfire Assessment Using Machine Learning Algorithms in Different Regions. *Fire Ecol.* **2024**, *20*, 104. <https://doi.org/10.1186/s42408-024-00335-2>.
24. Caiado, J.; Marques, M. Predicting Wildfire Occurrences in Portugal Using Machine Learning Classification Models. *Ecol. Inform.* **2025**, *92*, 103455. <https://doi.org/10.1016/j.ecoinf.2025.103455>.
25. Ali, S.Y.; Sedqi Kareem, O. Forest Fire Prediction Using Random Forest. *Eng. Technol. J.* **2025**, *10*. <https://doi.org/10.47191/etj/v10i05.44>.
26. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>.
27. Unik, M.; Sukaesih Sitanggang, I.; Syaufina, L.; Surati Jaya, I.N. Application of Random Forest Algorithm to Analyze the Confidence Level of Forest Fire Hotspots in Riau Peatland. *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (J. Nat. Resour. Environ. Manag.)* **2025**, *15*, 255. <https://doi.org/10.29244/jpsl.15.2.255>.
28. Yu, Y.; Mao, J.; Thornton, P.E.; Notaro, M.; Wulschleger, S.D.; Shi, X.; Hoffman, F.M.; Wang, Y. Quantifying the Drivers and Predictability of Seasonal Changes in African Fire. *Nat. Commun.* **2020**, *11*, 2893. <https://doi.org/10.1038/s41467-020-16692-w>.
29. Venevsky, S.; Le Page, Y.; Pereira, J.M.C.; Wu, C. Analysis Fire Patterns and Drivers with a Global SEVER-FIRE v1.0 Model Incorporated into Dynamic Global Vegetation Model and Satellite and on-Ground Observations. *Geosci. Model Dev.* **2019**, *12*, 89–110. <https://doi.org/10.5194/gmd-12-89-2019>.
30. Pye, J.M.; Rauscher, H.M.; Sands, Y.; Lee, D.C.; Beatty, J.S. *Advances in Threat Assessment and Their Application to Forest and Rangeland Management—Volume 1 and Volume 2*; U.S. Department of Agriculture, Forest Service, Pacific: Portland, OR, USA, 2010.
31. Vilar, L.; Camia, A.; San-Miguel-Ayanz, J.; Martín, M.P. Modeling Temporal Changes in Human-Caused Wildfires in Mediterranean Europe Based on Land Use-Land Cover Interfaces. *For. Ecol. Manag.* **2016**, *378*, 68–78. <https://doi.org/10.1016/j.foreco.2016.07.020>.
32. The United Republic of Tanzania (URT) *The 2022 Population and Housing Census: Iringa Region Basic Demographic and Socio-Economic Profile Report*; The United Republic of Tanzania (URT): Dodoma, Tanzania, 2024;.
33. Abdallah, J.M.; Monela, G.G. *Working Papers of the Finnish Forest Research Institute 50 MITMIOMBO-Management of Indigenous Tree Species for Ecosystem Restoration and Wood Production in Semi Overview of Miombo Woodlands in Tanzania*; (Working Papers of the Finnish Forest Research Institute 50); Finnish Forest Research Institute: Vantaa, Finland, 2007.
34. Giglio, L.; Schroeder, W.; Hall, J.V.; Justice, C.O. *MODIS Collection 6 Active Fire Product User's Guide Revision B*; NASA: Washington, DC, USA, 2018;.
35. Sulla-Menashe, D.; Friedl, M.A. User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product. 2022. Available online: https://lpdaac.usgs.gov/documents/101/MCD12_User_Guide_V6.pdf (accessed on 5 July 2026).
36. Morisette, J.T.; Giglio, L.; Csiszar, I.; Setzer, A.; Schroeder, W.; Morton, D.; Justice, C.O. Validation of MODIS Active Fire Detection Products Derived from Two Algorithms. *Earth Interact.* **2005**, *9*, 1–25. <https://doi.org/10.1175/EI141.1>.
37. Giglio, L.; Schroeder, W.; Justice, C.O. The Collection 6 MODIS Active Fire Detection Algorithm and Fire Products. *Remote Sens. Environ.* **2016**, *178*, 31–41. <https://doi.org/10.1016/j.rse.2016.02.054>.
38. Fusco, E.J.; Finn, J.T.; Abatzoglou, J.T.; Balch, J.K.; Dadashi, S.; Bradley, B.A. Detection Rates and Biases of Fire Observations from MODIS and Agency Reports in the Conterminous United States. *Remote Sens. Environ.* **2019**, *220*, 30–40. <https://doi.org/10.1016/j.rse.2018.10.028>.
39. Xie, L.; Zhang, R.; Zhan, J.; Li, S.; Shama, A.; Zhan, R.; Wang, T.; Lv, J.; Bao, X.; Wu, R. Wildfire Risk Assessment in Liangshan Prefecture, China Based on An Integration Machine Learning Algorithm. *Remote Sens.* **2022**, *14*, 4592. <https://doi.org/10.3390/rs14184592>.
40. Pan, J.; Wang, W.; Li, J. Building Probabilistic Models of Fire Occurrence and Fire Risk Zoning Using Logistic Regression in Shanxi Province, China. *Nat. Hazards* **2016**, *81*, 1879–1899. <https://doi.org/10.1007/s11069-016-2160-0>.
41. Quan, X.; Jiao, M.; He, Z.; Jaafari, A.; Xie, Q.; Lai, X. Effects of Different Sampling Strategies for Unburned Label Selection in Machine Learning Modelling of Wildfire Occurrence Probability. *Int. J. Wildland Fire* **2023**, *32*, 561–575. <https://doi.org/10.1071/WF21149>.
42. Sparks, A. Nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. *J. Open Source Softw.* **2018**, *3*, 1035. <https://doi.org/10.21105/joss.01035>.
43. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*. <https://doi.org/10.1029/2005RG000183>.
44. Barrington-Leigh, C.; Millard-Ball, A. A High-Resolution Global Time Series of Street-Network Sprawl. *Environ. Plan. B Urban Anal. City Sci.* **2025**, *52*, 1525–1536. <https://doi.org/10.1177/23998083241306829>.

45. Lehner, B.; Verdin, K.; Jarvis, A. New Global Hydrography Derived From Spaceborne Elevation Data. *Eos Trans. Am. Geophys. Union* **2008**, *89*, 93–94. <https://doi.org/10.1029/2008EO100001>.
46. Tan, J.; Xie, X.; Zuo, J.; Xing, X.; Liu, B.; Xia, Q.; Zhang, Y. Coupling Random Forest and Inverse Distance Weighting to Generate Climate Surfaces of Precipitation and Temperature with Multiple-Covariates. *J. Hydrol.* **2021**, *598*, 126270. <https://doi.org/10.1016/j.jhydrol.2021.126270>.
47. Oliveira, S.; Oehler, F.; San-Miguel-Ayanz, J.; Camia, A.; Pereira, J.M.C. Modeling Spatial Patterns of Fire Occurrence in Mediterranean Europe Using Multiple Regression and Random Forest. *For. Ecol. Manag.* **2012**, *275*, 117–129. <https://doi.org/10.1016/j.foreco.2012.03.003>.
48. Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G.; Jarvis, A. Very High Resolution Interpolated Climate Surfaces for Global Land Areas. *Int. J. Climatol.* **2005**, *25*, 1965–1978. <https://doi.org/10.1002/joc.1276>.
49. Tien Bui, D.; Le, K.-T.; Nguyen, V.; Le, H.; Revhaug, I. Tropical Forest Fire Susceptibility Mapping at the Cat Ba National Park Area, Hai Phong City, Vietnam, Using GIS-Based Kernel Logistic Regression. *Remote Sens.* **2016**, *8*, 347. <https://doi.org/10.3390/rs8040347>.
50. Jaafari, A.; Zenner, E.K.; Panahi, M.; Shahabi, H. Hybrid Artificial Intelligence Models Based on a Neuro-Fuzzy System and Metaheuristic Optimization Algorithms for Spatial Prediction of Wildfire Probability. *Agric. For. Meteorol.* **2019**, *266–267*, 198–207. <https://doi.org/10.1016/j.agrformet.2018.12.015>.
51. Iban, M.C.; Sekertekin, A. Machine Learning Based Wildfire Susceptibility Mapping Using Remotely Sensed Fire Data and GIS: A Case Study of Adana and Mersin Provinces, Turkey. *Ecol. Inform.* **2022**, *69*, 101647. <https://doi.org/10.1016/j.ecoinf.2022.101647>.
52. Muhammad, S.; Nizami, S.M.; Jiang, Q.O.; Mehmood, K.; Hussain, M.; Anees, S.A.; Shahzad, F.; Khan, W.R. Machine Learning-Based Forest Fire Vulnerability Assessment in Subtropical Chir Pine Forests of Pakistan. *Fire Ecol.* **2025**, *21*, 72. <https://doi.org/10.1186/s42408-025-00408-w>.
53. Radhi, A.A.; Ibrahim, A.A. An Intelligent IoT–Machine Learning Framework for Wildfire Detection and Prediction Using a Hybrid RF–XGB Model. *Sci. Rep.* **2026**, *16*, 19343. <https://doi.org/10.1038/s41598-026-52395-w>.
54. Martin-Gomez, V.; Von Ruetten, J.; Chiva, B.; Senande-Rivera, M.; Pinilla, M.; Burgues, J.; Baladima, F. Machine Learning-Based Wildfire Susceptibility Modeling for Catalonia the EGU General Assembly 2026, Vienna, Austria, 3–8 May 2026.
55. Bouargal, Y.; Tanarhte, M.; Stour, L.; Lafif, M. Wildfire Risk Assessment in Arid Oasis Ecosystems: An Integrated Machine Learning and Vulnerability Analysis Approach in Morocco. 2026. Available online: <https://meetingorganizer.copernicus.org/EGU26/EGU26-1544.html> (accessed on 5 July 2026).
56. Altmann, A.; Tološi, L.; Sander, O.; Lengauer, T. Permutation Importance: A Corrected Feature Importance Measure. *Bioinformatics* **2010**, *26*, 1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>.
57. Nurdianti, S.; Nugrahani, E.H.; Bukhari, F.; Najib, M.K.; Sebastian, D.M.; Nur Fallahi, P.A. Sensitivity and Feature Importance of Climate Factors and Evaluation of Different Machine Learning Models for Predicting Fire Hotspots in Kalimantan, Indonesia. *IAES Int. J. Artif. Intell. (IJ-AI)* **2024**, *13*, 2212. <https://doi.org/10.11591/ijai.v13.i2.pp2212-2225>.
58. Deng, J.; Hong, B.; Wang, W.; Gu, G. Daily Wildfire Risk Prediction by Mining Global and Local Spatio-Temporal Dependency. *Earth Sci. Inform.* **2025**, *18*, 316. <https://doi.org/10.1007/s12145-024-01652-5>.
59. O'Brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **2007**, *41*, 673–690. <https://doi.org/10.1007/s11135-006-9018-6>.
60. Rodrigues, M.; de la Riva, J. An Insight into Machine-Learning Algorithms to Model Human-Caused Wildfire Occurrence. *Environ. Model. Softw.* **2014**, *57*, 192–201. <https://doi.org/10.1016/j.envsoft.2014.03.003>.
61. Govil, K.; Welch, M.L.; Ball, J.T.; Pennypacker, C.R. Preliminary Results from a Wildfire Detection System Using Deep Learning on Remote Camera Images. *Remote Sens.* **2020**, *12*, 166. <https://doi.org/10.3390/rs12010166>.
62. Makumbura, R.K.; Dissanayake, P.; Gunathilake, M.B.; Rathnayake, N.; Kantamaneni, K.; Rathnayake, U. Spatial Mapping and Analysis of Forest Fire Risk Areas in Sri Lanka—Understanding Environmental Significance. *Case Stud. Chem. Environ. Eng.* **2024**, *9*, 100680. <https://doi.org/10.1016/j.csee.2024.100680>.
63. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinform.* **2007**, *8*, 25. <https://doi.org/10.1186/1471-2105-8-25>.
64. Tende, I.G. Prediction of Forest Fire Danger by Using BiGRU Deep Learning Model and Comparable Data Scaling Methods: A Case of SAO Hill Forest Plantation, Tanzania. *East Afr. J. Inf. Technol.* **2025**, *8*, 196–211. <https://doi.org/10.37284/eaajt.8.1.3135>.
65. Mambile, C.; Kaijage, S.; Leo, J. Deep Learning Models for Enhanced Forest-Fire Prediction at Mount Kilimanjaro, Tanzania: Integrating Satellite Images, Weather Data and Human Activities Data. *Nat. Hazards Res.* **2025**, *5*, 335–347. <https://doi.org/10.1016/j.nhres.2024.12.001>.

66. Mkiwa, B.P.; Mauya, E.W.; Jonas, J.N.; Mbeyale, G.E. Integrating Geospatial Tools in Mapping Forest Fire Severity and Burned Areas in the Western Usambara Mountain Forests, Lushoto, Tanzania. *PLoS ONE* **2025**, *20*, e0311865. <https://doi.org/10.1371/journal.pone.0311865>.
67. Baltazary, I.S.; Malila, B.P.; Lyimo, P.J.; Shirima, D.D.; Kilawe, C.J.; Nzunda, E.F. Land Use and Land Cover Changes Spatial Relationships with Ecological Fires and Their Implications for Sustainable Management and Conservation of Tanzania's Miombo Woodlands. *Discov. For.* **2025**, *1*, 23. <https://doi.org/10.1007/s44415-025-00018-z>.
68. Katani, J.Z.; Madoffe, S.S.; Amanzi, N.S.; Rija, A.A.; Midtgaard, F.; Mbeyale, G.; Zahabu, E.; Tarimo, B.C. Assessment of Fire Prevalence and Reduction Strategies in Miombo Woodlands of Eastern Tanzania. *Tanzan. J. For. Nat. Conserv.* **2014**, *84*, 1–7. <https://doi.org/10.4314/tjfn.v84i1>.
69. Shivashankar, S.K.; Prajwal, M.D.; Likith Raj, K.R.; Tanya Priyadarshini, A.R.; Manvitha, S.M. Forest Fire Prediction Using Random Forest Regressor: A Comprehensive Machine Learning Approach. *Int. J. Innov. Sci. Res. Technol. (IJISRT)* **2024**, 2063–2071. <https://doi.org/10.38124/ijisrt/ijisrt24sep1290>.
70. Yang, C.; Yao, P.; Wang, Q.; Wang, S.; Xing, D.; Wang, Y.; Zhang, J. XGBoost-Based Susceptibility Model Exhibits High Accuracy and Robustness in Plateau Forest Fire Prediction. *Forests* **2026**, *17*, 74. <https://doi.org/10.3390/f17010074>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.