

Explainable random forest modelling of recorded wildfire occurrence using multi-source geospatial data in the Aurès Mountains of Algeria

Razika Sahnouni^{1,2*} , Asma Kahli³, Ichrak Bahou¹, Nada Kraimia¹

¹ Higher National School of Forests, Khenchela 40000, Algeria

² Laboratory of Algerian Forests and Climate Change, Higher National School of Forests, Khenchela 40000, Algeria

³ Laboratory of Agriculture and Ecosystem Functioning, Department of Agronomy, Faculty of Nature and Life Sciences, University of Chadli Bendjedid, El Tarf 36000, Algeria

* Corresponding author's e-mail: sahnouni.razika@ensf.dz

ABSTRACT

This study develops an explainable machine-learning framework to investigate recorded wildfire occurrence in the Aurès Mountains of north-eastern Algeria using multi-source geospatial data. A total of 584 wildfire occurrence locations extracted from NASA's Fire Information for Resource Management System (FIRMS) for 2015–2025 were combined with 1.926 randomly generated background locations, yielding 2.510 observations. Ten explanatory variables representing vegetation condition, climate, topography, and anthropogenic accessibility were derived from Sentinel-2 imagery, WorldClim, the USGS Digital Elevation Model, and OpenStreetMap. A Random Forest model was developed following hyperparameter optimisation. Performance was evaluated using conventional random train–test partitioning and five-fold stratified cross-validation, complemented by five-fold spatial-block cross-validation using 10×10 km blocks to assess geographic generalisation. Under random evaluation, the model achieved an accuracy of 90.84% and an ROC–AUC of 0.939, with a mean five-fold ROC–AUC of 0.937 ± 0.014 . Spatial-block cross-validation produced a substantially lower mean ROC–AUC of 0.714 ± 0.111 , indicating reduced performance across spatially separated areas. Altitude and temperature were consistently identified among the leading predictors, while SHAP analysis indicated that higher temperature and NDVI generally contributed to higher modelled likelihood of recorded wildfire occurrence, whereas higher precipitation and NDMI contributed to lower likelihood. These findings highlight the importance of climatic, vegetation, topographic, and accessibility-related spatial gradients. However, predictor collinearity and the use of background locations rather than confirmed fire-absence observations require cautious interpretation. The framework provides an interpretable basis for wildfire susceptibility assessment and spatially targeted prevention and forest management, while geographic transferability requires further evaluation.

Keywords: wildfire occurrence, Random Forest, geospatial data, explainable machine learning, Aurès Mountains, Algeria.

INTRODUCTION

Wildfires are among the most serious environmental disturbances affecting terrestrial ecosystems, causing biodiversity loss, ecosystem degradation, disruption of carbon cycling, and substantial socio-economic impacts (IPCC, 2023). Climate change is increasing

fire-conductive conditions through rising temperatures, prolonged droughts, reduced fuel moisture, and more frequent extreme weather events, with particularly strong effects in Mediterranean ecosystems characterised by hot, dry summers and flammable vegetation (Curt et al., 2020; Ruffault et al., 2018). However, wildfire occurrence is not controlled by climate alone. Spatial patterns of

fire occurrence reflect the combined influence of climatic, vegetation, topographic, and anthropogenic factors, whose relative contributions vary across landscapes (Moreira et al., 2020; Ruffault et al., 2020). These relationships are often non-linear and spatially heterogeneous, while correlations among environmental predictors can complicate their interpretation.

Remote sensing and geographic information systems (GIS) provide increasingly effective means of characterising the environmental and anthropogenic conditions associated with wildfire occurrence. Satellite observations such as Sentinel-2 and Landsat can be integrated with climatic, topographic, and accessibility variables to support spatial wildfire modelling (Chuvieco et al., 2020; Mallinis et al., 2023). Among machine-learning approaches, Random Forest (RF) is widely used in environmental applications because of its ability to capture nonlinear relationships and interactions among predictors while remaining relatively robust to complex datasets (Belgiu and Drăguț, 2016). RF has consequently been applied extensively to wildfire susceptibility and occurrence modelling in Mediterranean and other fire-prone regions (Costafreda-Aumedes et al., 2017; Trucchia et al., 2022). However, for spatially structured environmental data, model performance depends not only on predictive modelling but also on appropriate sampling, validation, and interpretation strategies.

An important limitation of machine-learning models is that their predictions are not always readily interpretable. Explainable artificial intelligence (XAI) methods can provide complementary information about predictor contributions and improve model transparency. Gini-based feature importance and Permutation Feature Importance (PFI) quantify the relative contribution of predictors, whereas SHAP (SHapley Additive exPlanations) provides information on the magnitude and direction of individual predictor effects on model predictions (Lundberg and Lee, 2017; Lundberg et al., 2020). However, correlated predictors may share predictive information, meaning that feature importance should not be interpreted as evidence of independent causal effects (Dormann et al., 2013; Fisher et al., 2019). Therefore, XAI results need to be considered within the ecological and statistical structure of the predictor set.

The Mediterranean Basin is particularly vulnerable to wildfire because of its climatic conditions, flammable vegetation, complex terrain,

and increasing anthropogenic pressure. Climate change is expected to further intensify fire-conducive conditions through increasing drought and heatwave frequency (Turco et al., 2018; IPCC, 2023). In Algeria, these pressures are especially relevant in mountainous environments, where climatic gradients, topography, vegetation characteristics, and human accessibility may jointly influence wildfire occurrence. Despite increasing research on Mediterranean wildfires, spatially explicit assessments remain comparatively limited for North African mountain ecosystems. The Aurès Mountains constitute a heterogeneous Mediterranean–semi-arid landscape characterised by strong altitudinal and environmental gradients, making them an important setting for investigating the spatial factors associated with recorded wildfire occurrence.

Wildfire datasets derived from satellite active-fire detections also require careful interpretation. A recorded detection provides evidence that a fire was detected, whereas the absence of a detection does not necessarily indicate confirmed fire absence. Consequently, background locations should be interpreted as samples of available environmental conditions rather than confirmed non-fire locations. Furthermore, spatial clustering and autocorrelation may lead to overly optimistic model performance when neighbouring observations are randomly divided between training and evaluation datasets. Spatial-block cross-validation can provide a more demanding assessment of geographic generalisation by reducing spatial overlap between training and validation observations.

Accordingly, this study investigates the spatial associations between recorded wildfire occurrence and climatic, vegetation, topographic, and anthropogenic gradients in the Aurès Mountains of north-eastern Algeria using an explainable Random Forest framework and multi-source geospatial data. Specifically, the study aims to: (i) characterise the spatial distribution of recorded wildfire occurrences using multi-source geospatial predictors; (ii) evaluate RF performance and geographic generalisation using conventional random evaluation and spatial-block cross-validation; (iii) identify the predictors most strongly associated with wildfire occurrence using Gini importance, permutation feature importance, and SHAP; and (iv) examine predictor correlations and their implications for interpreting modelled feature contributions. The results are interpreted as spatial associations rather than causal relationships or temporal

forecasts, providing an evidence-based basis for wildfire susceptibility assessment, prevention, surveillance, and forest-management planning in Mediterranean mountain ecosystems.

STUDY AREA

The study area is situated in the Aurès Mountains of north-eastern Algeria (Figure 1). It extends between 6.00° and 7.11° E and 34.99° and 35.51° N, covering an area of approximately 6,000 km². With elevations exceeding 2,000 m, the Aurès massif represents a heterogeneous Mediterranean mountain environment located at the transition between the Tellian Atlas and the pre-Saharan regions (Ballais and Cohen, 1981). The massif encompasses several municipalities within the wilayas of Batna and Khenchela and includes important forest formations dominated by Aleppo pine (*Pinus halepensis*), Atlas cedar (*Cedrus atlantica*), holm oak (*Quercus ilex*), and junipers (*Juniperus* spp.) (Beghami et al., 2012).

The study area is characterised by pronounced topographic variability, including steep slopes,

dissected valleys, and substantial differences in elevation. This environmental heterogeneity is accompanied by spatial variations in temperature, precipitation, and vegetation conditions that may influence the spatial distribution of recorded wildfire occurrence. According to WorldClim Version 2.1, the region has a semi-arid Mediterranean climate, characterised by hot, dry summers and cold winters. The mean maximum temperature is approximately 33.9 °C in July, whereas the mean minimum temperature is approximately −0.05 °C in January. Average annual precipitation is approximately 346 mm, with most rainfall occurring between autumn and spring (Fick and Hijmans, 2017).

Fire represents an important disturbance affecting forest ecosystems in the region. According to Global Forest Watch, between 2001 and 2024, fire was associated with approximately 83% of tree-cover loss in the wilaya of Khenchela and 68% in the wilaya of Batna (Global Forest Watch, 2025). These figures highlight the substantial contribution of fire-associated disturbance to recent tree-cover loss in the region and provide an important context for assessing the spatial distribution of recorded wildfire occurrence.

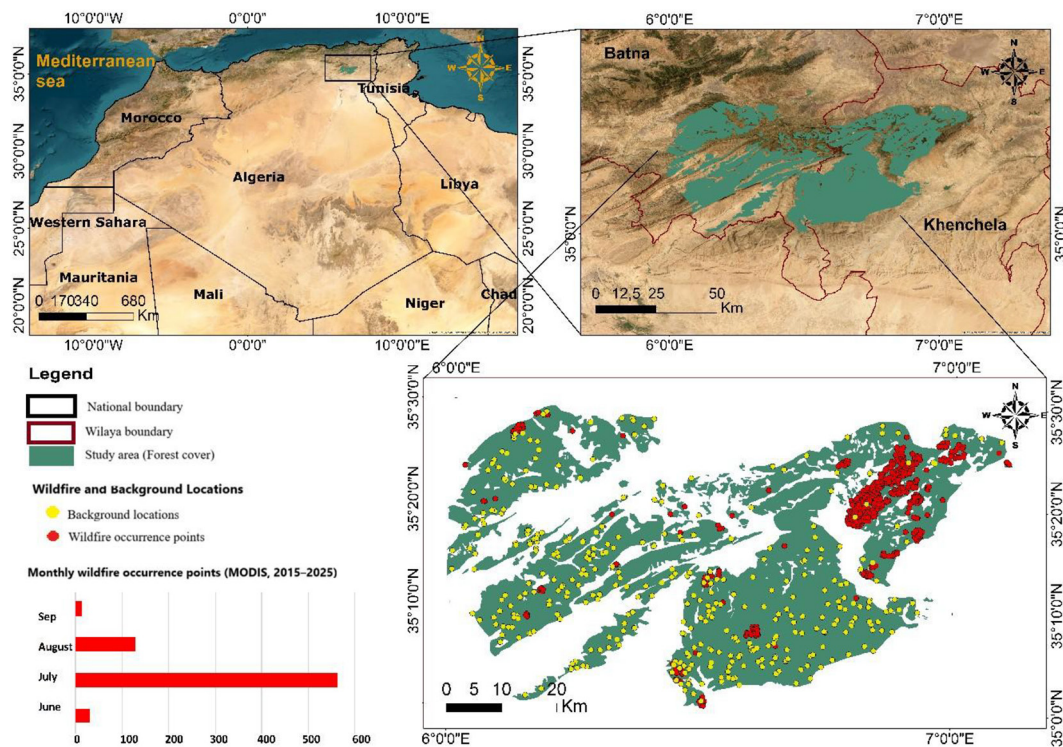


Figure 1. Geographic location of the study area and spatial distribution of recorded wildfire occurrences and background locations. Wildfire occurrence points were derived from NASA's Fire Information for Resource Management System (FIRMS) MODIS active-fire detections for June–September during 2015–2025. Background locations were randomly generated within the forested area. The lower-left panel shows the monthly distribution of MODIS active-fire detections during the study period

RESEARCH METHODOLOGY

Data collection and preparation

Wildfire inventory (response variable)

The wildfire inventory was compiled from NASA's Fire Information for Resource Management System (FIRMS). Active-fire detections from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors aboard the Terra and Aqua satellites were used to represent recorded wildfire occurrences during the main fire season (June–September) from 2015 to 2025. MODIS Collection 6 provides active-fire detections at an approximate nominal spatial resolution of 1 km and has been widely applied for wildfire monitoring in Mediterranean ecosystems (Chuvieco et al., 2019). The initial FIRMS dataset underwent quality control to remove duplicate detections and records located outside the forested areas of the study region. After filtering, 584 recorded wildfire occurrence locations were retained for subsequent modelling. Because the absence of a FIRMS detection cannot be interpreted as confirmed fire absence, locations without recorded wildfire detections were treated as background locations rather than true-absence observations. Accordingly, 1,926 background points were randomly generated within the forested area using ArcGIS 10.8, independently of the environmental predictors and prior to model fitting. Background points coinciding with recorded wildfire occurrence locations were excluded. The final modelling dataset comprised 2,510 locations, including 584 recorded wildfire occurrence locations (23.27%) and 1,926 background locations (76.73%). The response variable was coded as 1 for recorded wildfire occurrence and 0 for sampled background locations. Since FIRMS detections represent satellite-observed active fires rather than independently delineated wildfire events, model outputs were interpreted as predictive scores representing the discrimination of recorded wildfire occurrence locations from sampled background locations, rather than as definitive fire/no-fire probabilities or causal relationships. The term “recorded wildfire occurrence” is therefore used throughout the study to reflect the nature and limitations of the response data.

Independent variables

Ten explanatory variables were selected to characterise environmental and anthropogenic

conditions potentially associated with the spatial distribution of recorded wildfire occurrences. The variables comprised vegetation, climate, topographic, and anthropogenic factors (Curt et al., 2016; Table 1). Their selection was guided by previous studies of wildfire occurrence in Mediterranean forest ecosystems and by the environmental characteristics of the study area.

Vegetation conditions were represented by the normalised difference vegetation index (NDVI) and normalised difference moisture index (NDMI), derived from atmospherically corrected Sentinel-2 Level-2A surface reflectance data. Processing was restricted to the main fire season (June–September) over the 2015–2025 study period. Cloud, cirrus, cloud-shadow, and snow contamination were identified using the Sentinel-2 Scene Classification Layer (SCL), and scenes with more than 20% cloud cover over the study area were excluded. Following pixel-level masking, seasonal composites were generated from the remaining valid observations, and the mean of all available June–September observations across 2015–2025 was calculated for each pixel. The resulting layers represent the multi-year mean spatial patterns of vegetation greenness and moisture conditions during the main fire season.

NDVI was calculated using the near-infrared (NIR; B8) and red (B4) bands, while NDMI was derived from the NIR (B8) and short-wave infrared (SWIR; B11) bands, according to Equations 1 and 2, respectively. B4 and B8 have a native spatial resolution of 10 m, whereas B11 has a native spatial resolution of 20 m. For NDMI calculation, B8 was resampled to the 20-m spatial resolution of B11, and the resulting vegetation layers were subsequently harmonised to the common 20-m modelling grid used for predictor extraction.

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (\text{Rouse et al., 1973}) \quad (1)$$

$$NDMI = \frac{NIR-SWIR}{NIR+SWIR} \quad (\text{Gao, 1996}) \quad (2)$$

where: *NIR* denotes near-infrared reflectance from Sentinel-2 Band 8), *Red* denotes red reflectance from Band 4, and *SWIR* denotes short-wave infrared reflectance from Band 11.

Climatic conditions were represented by mean air temperature and precipitation obtained from the WorldClim Version 2.1 database at an approximately 1-km spatial resolution (Fick and Hijmans,

Table 1. Predictor variables used in the Random Forest model of recorded wildfire occurrence

Category	Predictor	Data source	Native spatial resolution	Processing / derivation
Vegetation	NDVI	Sentinel-2 Level-2A surface reflectance	10 m (B4, B8)	June–September multi-year mean, 2015–2025; cloud, cirrus, cloud-shadow, and snow masking using the Scene Classification Layer
Vegetation	NDMI	Sentinel-2 Level-2A surface reflectance	20 m (B11; B8 resampled as required)	June–September multi-year mean, 2015–2025; cloud, cirrus, cloud-shadow, and snow masking using the Scene Classification Layer
Climate	Air temperature	WorldClim Version 2.1	~1 km	Mean June–September climatic conditions representing the broad fire-season spatial gradient
Climate	Precipitation	WorldClim Version 2.1	~1 km	Mean June–September climatic conditions representing the broad fire-season spatial gradient
Topography	Elevation	USGS DEM	30 m	Extracted directly from the DEM
Topography	Slope	USGS DEM	30 m	Derived from the DEM using ArcGIS Spatial Analyst
Topography	Aspect	USGS DEM	30 m	Derived from the DEM using ArcGIS Spatial Analyst
Topography	Profile curvature	USGS DEM	30 m	Derived from the DEM using ArcGIS Spatial Analyst
Anthropogenic	Distance to roads	OpenStreetMap; checked against high-resolution satellite imagery	Vector	Euclidean distance calculated in ArcGIS 10.8
Anthropogenic	Distance to residential areas	OpenStreetMap; checked against high-resolution satellite imagery	Vector	Euclidean distance calculated in ArcGIS 10.8

2017). June–September mean temperature and precipitation were calculated from the corresponding WorldClim monthly layers to characterise the broad-scale spatial climatic gradient during the main fire season. Given the substantially coarser spatial resolution of WorldClim relative to the final modelling grid, these variables were considered indicators of broad-scale climatic variability rather than fine-scale meteorological conditions immediately preceding individual wildfire events.

Four topographic variables—elevation, slope, aspect, and profile curvature—were derived from the USGS 30-m digital elevation model (DEM) to represent spatial variation in terrain conditions potentially associated with microclimate, solar exposure, moisture availability, vegetation distribution, and fire behaviour. These relatively stable terrain characteristics were treated as time-invariant predictors (Rossa et al., 2015). Because elevation was strongly correlated with temperature and precipitation, its importance was interpreted as representing, at least partly, a broader environmental gradient shared with these climatic variables rather than as an independent effect of elevation.

Anthropogenic influence was represented by distance to roads and distance to residential areas. Road and residential-area data were obtained

from OpenStreetMap and cross-checked against high-resolution satellite imagery to improve the spatial representation of mapped infrastructure. Euclidean distance surfaces were generated in ArcGIS 10.8 for both variables. These predictors were used to characterise spatial accessibility and proximity to human infrastructure that may be associated with recorded wildfire occurrence (Oliveira et al., 2012; Curt et al., 2016). These variables were interpreted as spatial proxies for accessibility and human presence rather than as direct measures of human-caused ignition.

All explanatory variables were subsequently harmonised to a common spatial reference system and modelling grid before extraction at wildfire occurrence and background locations. Given differences in spatial resolution, temporal representation, and correlation among predictors, model-derived contributions were interpreted in relation to the overall environmental structure of the study area rather than as independent causal effects.

Exploratory data analysis

Prior to model development, exploratory data analysis (EDA) was conducted to characterise the predictor distributions and assess relationships

among explanatory variables. For each quantitative predictor, minimum, maximum, mean, median, standard deviation, and coefficient of variation (CV) were calculated. CV was not reported for aspect, because it is a circular variable, or for NDMI and profile curvature, because their means were close to zero, making this relative measure of variability inappropriate. Skewness and excess kurtosis were calculated to further describe distributional shape, while histograms and kernel density estimates (KDEs) were used to visualise the distributions and identify potential extreme values. The proportions of recorded wildfire occurrence and background locations were also examined to characterise the response dataset.

Pairwise relationships among predictors were assessed using Pearson's correlation coefficient, with $|r| \geq 0.80$ considered indicative of strong linear association (Dormann et al., 2013; James et al., 2021). Although Random Forest is relatively robust to multicollinearity compared with conventional regression models, correlated predictors can influence the distribution and interpretation of variable-importance measures (Breiman, 2001). Therefore, correlation patterns were considered when interpreting Gini-based importance, permutation feature importance (PFI), and SHAP values, and predictor importance was interpreted as model-dependent contribution rather than as an independent or causal effect (Molnar, 2022).

Random forest modelling

Data preparation and model evaluation

Following preprocessing and spatial harmonisation, the final modelling dataset comprised 2,510 observations, including 584 recorded wildfire occurrence locations and 1,926 background locations, each characterised by ten explanatory variables. Because the predictors represent multi-year spatial conditions rather than environmental conditions immediately preceding individual fire events, the modelling framework was designed to evaluate spatial associations between recorded wildfire occurrence and environmental characteristics across the study area during 2015–2025.

For conventional model evaluation, the complete dataset was randomly divided into a stratified 80% training set ($n = 2,008$) and a 20% independent test set ($n = 502$), preserving the relative proportions of recorded wildfire occurrence and background observations in both subsets. The

training set was used for model development, including hyperparameter optimisation and internal cross-validation, whereas the test set was withheld from model selection and used exclusively for final performance evaluation. A fixed random seed was applied to ensure reproducibility.

To assess model generalisation under spatial separation, a complementary spatial-block cross-validation scheme was implemented. The study area was divided into 10×10 km blocks, and five-fold stratified group cross-validation was performed using the blocks as grouping units. All observations within a block were assigned to the same fold, preventing locations from the same block from being distributed across training and validation subsets. This procedure reduced the influence of local spatial dependence and provided a more stringent assessment of geographic generalisation. Spatial blocking was treated as a strategy for reducing spatial dependence rather than as a guarantee of complete spatial independence.

Hyperparameter optimisation

The RF model was implemented using the Scikit-learn library (Pedregosa et al., 2011). Hyperparameters were optimised using *RandomizedSearchCV*, which evaluated 30 randomly selected combinations of candidate settings ($n_iter = 30$). The search space included the number of trees ($n_estimators$), maximum tree depth (max_depth), minimum samples required for node splitting ($min_samples_split$), minimum samples per terminal leaf ($min_samples_leaf$), number of predictors considered at each split ($max_features$), bootstrap resampling, and class weighting. Class weighting was included among the candidate hyperparameter settings to reduce the influence of the unequal class proportions during model fitting.

Hyperparameter optimisation was performed using five-fold stratified cross-validation within the training set, with ROC-AUC used as the optimisation metric. The independent test set was excluded from model selection and used only for final performance evaluation. The best-performing configuration was subsequently refitted to the complete training set and evaluated once on the independent test set. A fixed random state was used throughout the optimisation, model fitting, and data partitioning procedures to ensure reproducibility.

The selected RF configuration comprised 500 trees ($n_estimators = 500$), a maximum tree depth of 40 ($max_depth = 40$), $min_samples_split = 2$, $min_samples_leaf = 2$, $max_features = sqrt$, and no bootstrap resampling ($bootstrap = False$), with class weighting set to balanced.

Explainable artificial intelligence

To enhance the interpretability of the RF model, three complementary explainable artificial intelligence (XAI) methods were applied: Gini-based feature importance, permutation feature importance (PFI), and SHapley Additive exPlanations (SHAP) (Molnar, 2022). These approaches capture different aspects of predictor importance: Gini importance is derived from the internal structure of the RF, PFI assesses the change in predictive performance following permutation of a predictor, and SHAP quantifies its contribution to individual model predictions.

Gini-based feature importance

Gini importance was extracted directly from the fitted RF and calculated as the mean reduction in Gini impurity attributable to splits involving each predictor across the decision trees (Breiman, 2001). Higher values indicate greater contribution to node-purity reduction. Because impurity-based importance can be affected by predictor characteristics and shared information among correlated variables, it was interpreted jointly with PFI and SHAP rather than considered as an independent measure of predictor effect.

Permutation feature importance (PFI)

PFI was calculated on the independent test set after final model fitting. For each predictor, its values were randomly permuted while the remaining predictors were unchanged, and the resulting change in predictive performance was recorded. The permutation procedure was repeated 30 times per predictor, and the mean importance and standard deviation were calculated. A larger decrease in predictive performance following permutation indicated greater reliance of the fitted model on the information provided by that predictor (Fisher et al., 2019; Molnar, 2022). Because correlated predictors can retain overlapping information after permutation, PFI may underestimate the contribution of individual predictors within correlated groups.

SHapley additive exPlanations (SHAP)

SHAP was used to quantify the contribution of individual predictors to RF predictions based on Shapley values from cooperative game theory (Lundberg and Lee, 2017). Values were estimated using TreeExplainer, which is designed for efficient interpretation of tree-based models (Lundberg et al., 2020). SHAP analysis was conducted on the independent test set. Global importance was quantified using the mean absolute SHAP value, with larger values indicating greater overall contribution to model predictions. The signed SHAP values in the SHAP summary plots were subsequently examined to assess whether higher or lower predictor values were associated with increases or decreases in the modelled recorded-wildfire occurrence score.

The three XAI approaches were subsequently compared with the predictor correlation structure to identify consistent patterns of importance and potential differences among methods. In particular, strong correlations among vegetation, climatic, and topographic variables were considered when interpreting individual predictor contributions, as importance attributed to one variable may partly reflect information shared with correlated predictors.

RESULTS AND DISCUSSION

Relative importance of predictors associated with recorded wildfire occurrence

The integrated XAI analysis showed that discrimination between recorded wildfire occurrence locations and background locations was associated with a combination of topographic, climatic, vegetation, and anthropogenic predictors. Altitude and temperature consistently ranked among the most important predictors across the three XAI approaches, although the relative ranking of the remaining predictors varied among methods. NDVI ranked third according to Gini importance and mean absolute SHAP, but sixth according to PFI, whereas distance to residential areas ranked third according to PFI but sixth according to mean absolute SHAP. Precipitation and distance to roads also showed moderate-to-high predictive contributions.

The differences among importance measures reflect both their distinct definitions of predictor

contribution and the shared information among correlated variables (Fisher et al., 2019). Altitude was strongly correlated with temperature and precipitation, while NDVI was strongly correlated with NDMI. Thus, the importance attributed to individual predictors may partly represent shared information rather than unique predictor effects. This is particularly relevant for altitude in the Aurès Mountains, where elevation is closely associated with climatic and ecological gradients, including temperature, precipitation, and vegetation distribution. Its high importance may therefore capture a broader environmental gradient rather than an isolated elevation effect. Similarly, the contribution of temperature should be interpreted in relation to its spatial association with elevation and other environmental variables.

The contrasting rankings of NDVI and distance to residential areas further illustrate the dependence of importance estimates on the XAI method. The lower PFI ranking of NDVI relative to its Gini and SHAP rankings may indicate that much of its predictive information was shared with other environmental variables. In contrast, the relatively high PFI contribution of distance to residential areas suggests that proximity to settlements provided useful predictive information despite its lower SHAP ranking. Overall, the agreement among methods for the leading predictors, together with differences in the ranking of secondary variables, indicates that recorded wildfire occurrence was associated with interacting environmental and anthropogenic gradients rather than a single dominant predictor. These

findings represent model-dependent predictive contributions and spatial associations and should not be interpreted as independent causal effects.

Exploratory data analysis

Descriptive statistics and distribution of predictors

The descriptive statistics of the ten environmental and anthropogenic predictors used for recorded wildfire occurrence modelling are presented in Table 2. Altitude ranged from 835.12 to 2,136.45 m (mean = 1,382.52 ± 197.52 m; CV = 14.29%), while slope ranged from 0 to 90° (mean = 23.60°; CV = 98.91%). Aspect covered the full angular range from 0 to 359.96°, reflecting substantial topographic heterogeneity across the study area. These topographic gradients may influence fuel moisture, vegetation distribution, and fire propagation through their effects on local environmental conditions (Pausas and Fernández-Muñoz, 2012; Ruffault et al., 2018; San-Miguel-Ayán et al., 2024).

Climatic variables showed comparatively limited spatial variability, with temperature ranging from 16.66 to 24.55 °C (CV = 6.93%) and seasonal precipitation from 80 to 167 mm (CV = 10.54%). These variables represent the broad climatic gradient during the summer fire season and may influence vegetation condition and fuel moisture (Ruffault et al., 2020). NDVI ranged from −0.059 to 0.473, whereas NDMI ranged from −0.162 to 0.232, indicating spatial variation in vegetation productivity and moisture status

Table 2. Descriptive statistics of the environmental and anthropogenic predictors used to model recorded wildfire occurrence

Variable	Minimum	Maximum	Mean	Median	Standard deviation	CV (%)
Altitude (m)	835.118	2,136.450	1,382.522	1,381.765	197.520	14.287
Profile curvature	−8.556	9.005	0.049	0.016	2.228	—
Aspect (°)	0.000	359.964	177.187	175.322	107.786	—
Distance to residential areas (m)	0.000	18,617.300	4,038.496	3,042.300	3,348.573	82.916
Distance to roads (m)	0.000	10,957.500	1,437.824	953.834	1,576.569	109.650
NDMI	−0.162	0.232	−0.023	−0.033	0.063	—
NDVI	−0.059	0.473	0.220	0.211	0.082	37.155
Slope (°)	0.000	90.000	23.601	17.454	23.344	98.914
Temperature (°C)	16.664	24.550	20.203	20.015	1.400	6.928
Precipitation (mm)	80.000	167.000	118.867	120.000	12.533	10.544

Note: CV = coefficient of variation. CV was not calculated for aspect because it is a circular variable and for profile curvature and NDMI because their means were close to zero, making the ratio of standard deviation to mean unstable and difficult to interpret.

(Mallinis et al., 2023). The accessibility variables showed the greatest relative dispersion, with CVs of 109.65% for distance to roads and 82.92% for distance to residential areas. These variables represent spatial proxies for accessibility and human presence that may contribute to anthropogenic wildfire patterns (Moreira et al., 2020; San-Miguel-Ayanz et al., 2024).

The distributional characteristics of the predictors are summarised in Table 3 and Figure 2. Most variables showed relatively low to moderate skewness and excess kurtosis. Distance to roads exhibited the strongest positive skewness (2.253) and highest excess kurtosis (6.563), followed by distance to residential areas (skewness = 1.663; excess kurtosis = 2.641). Slope also showed moderate positive skewness (0.863), whereas the remaining predictors displayed comparatively limited departures from symmetry. These distributional characteristics do not require transformation for Random Forest modelling because tree-based methods do not assume normally distributed predictors or homoscedastic residuals (Valavi et al., 2021).

Class composition of the modelling dataset

The final modelling dataset comprised 2,510 observations, including 584 recorded wildfire occurrence locations (23.27%) and 1,926 background locations (76.73%) (Table 4). The original class proportions were retained without oversampling or undersampling, while class weighting was incorporated into the Random Forest model to account for class imbalance. Background locations represent sampled locations without a recorded FIRMS detection and therefore do not constitute confirmed fire-absence observations. Accordingly, the classification task represents the spatial discrimination of recorded wildfire occurrence locations relative to sampled background conditions. The exploratory analysis revealed substantial spatial variability among the predictors, particularly in topography, vegetation condition, and accessibility. This heterogeneity provided the

Table 3. Distributional characteristics of the predictors used to model recorded wildfire occurrence

Variable	Skewness	Excess kurtosis
Altitude	0.142	0.075
Profile curvature	−0.030	0.868
Aspect	0.014	−1.180
Distance to residential areas	1.663	2.641
Distance to roads	2.253	6.563
NDMI	0.641	−0.033
NDVI	0.341	−0.636
Slope	0.863	−0.215
Temperature	0.597	0.258
Precipitation	−0.093	0.381

Note: Kurtosis is reported as excess kurtosis, for which a normal distribution has a value of 0.

environmental basis for examining the spatial predictors associated with recorded wildfire occurrence using the Random Forest framework.

Correlation analysis of predictor variables

Pearson correlation analysis was used to assess pairwise relationships among the environmental and anthropogenic predictors (Figure 3). Most predictor pairs showed weak to moderate associations, although several strong correlations were identified. The strongest positive correlation occurred between NDVI and NDMI ($r = 0.868$), indicating substantial overlap in the information represented by the two vegetation indices. Altitude was strongly positively correlated with precipitation ($r = 0.836$) and strongly negatively correlated with temperature ($r = -0.835$), while temperature and precipitation were also strongly negatively correlated ($r = -0.862$). These relationships reflect the pronounced environmental gradient across the Aurès Mountains, where elevation is associated with substantial variations in temperature, precipitation, and vegetation conditions.

A moderate positive correlation was observed between distance to roads and distance to

Table 4. Composition of the modelling dataset according to recorded wildfire occurrence and background status

Class	Number of observations	Percentage (%)
Background locations	1,926	76.73
Recorded wildfire occurrence locations	584	23.27
Total	2,510	100.00

Note: background locations represent sampled locations without a recorded FIRMS detection and should not be interpreted as confirmed fire-absence locations.

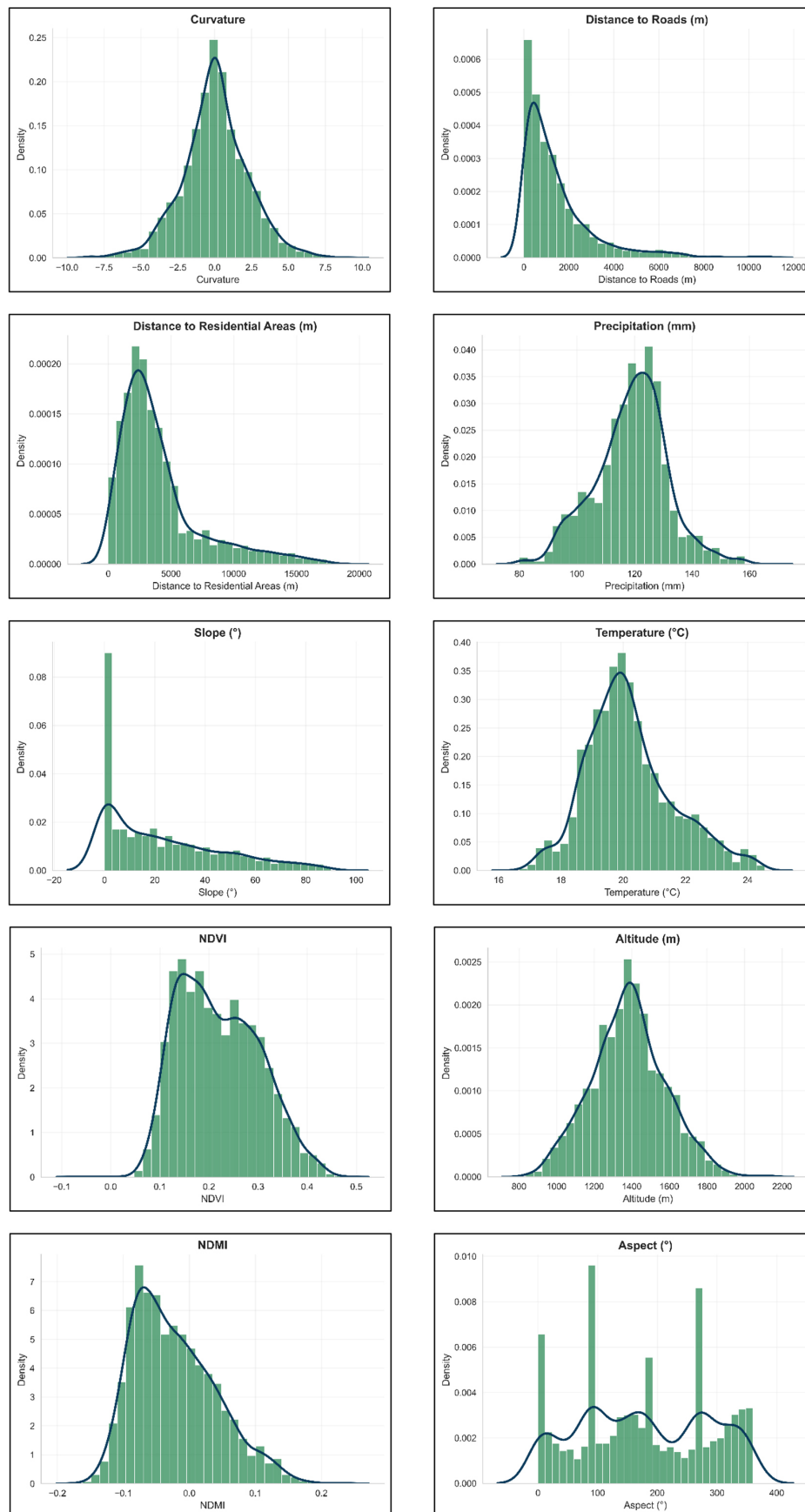


Figure 2. Distribution of the environmental and anthropogenic predictor variables used for wildfire occurrence modelling in the Aurès Mountains. Histograms and kernel density estimates (KDEs) illustrate the distributions of the topographic, climatic, vegetation, and accessibility-related variables included in the Random Forest model.

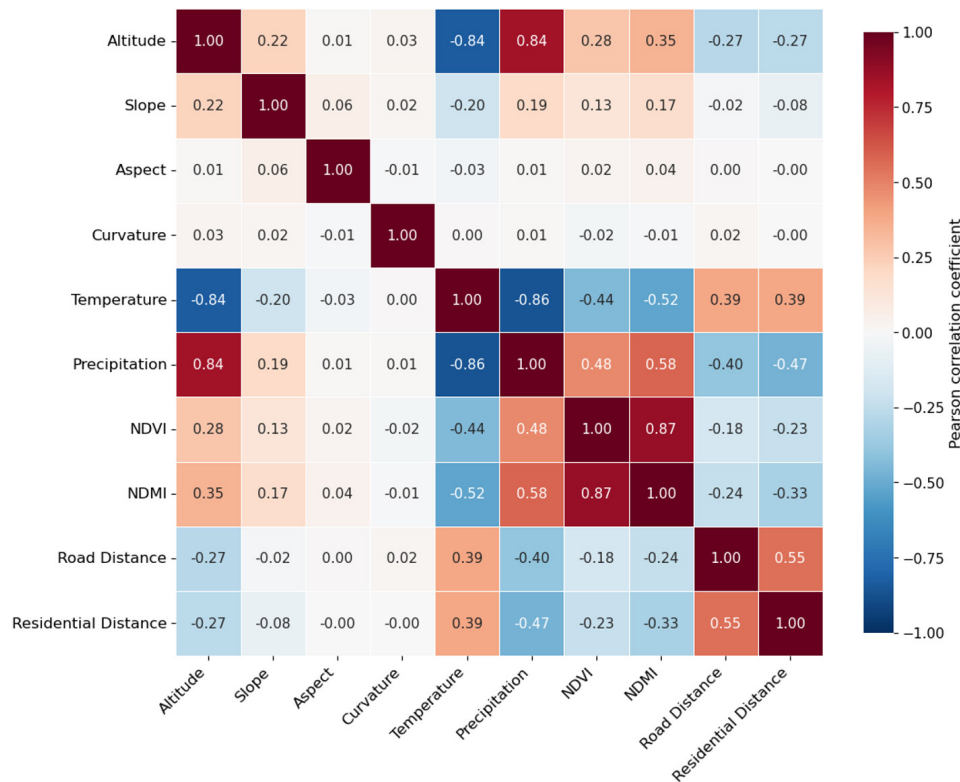


Figure 3. Pearson correlation matrix of the environmental and anthropogenic predictors used in the Random Forest model of recorded wildfire occurrence in the Aurès Mountains, Algeria. Strong correlations are observed between NDVI and NDMI and among altitude, temperature, and precipitation, highlighting shared information that was considered when interpreting predictor-importance and explainability results

residential areas ($r = 0.554$), suggesting partially overlapping spatial patterns associated with accessibility and human infrastructure. The remaining predictor pairs showed comparatively weak correlations.

No predictors were removed solely on the basis of pairwise correlation because Random Forest can accommodate correlated predictors without requiring statistical independence (Breiman, 2001). Nevertheless, the correlation structure was considered when interpreting predictor-importance measures. In particular, the strong associations among altitude, temperature, precipitation, NDVI, and NDMI indicate that some predictors contain shared information. Consequently, individual importance values were interpreted as model-dependent predictive contributions rather than independent effects. The correlated predictors were

therefore retained to preserve potentially relevant environmental information, while Gini importance, permutation feature importance (PFI), and SHAP values were interpreted jointly. Differences in predictor rankings among these approaches were considered in relation to the observed correlation structure. Overall, the correlation analysis indicates that the predictors represent interconnected environmental and anthropogenic gradients rather than statistically independent dimensions of recorded wildfire occurrence.

Random forest model performance

The Random Forest model was first evaluated using a stratified 80/20 random hold-out partition. The resulting confusion matrix is presented in Table 5. On the 502-observation independent

Table 5. Confusion matrix of the Random Forest model under stratified random hold-out validation

Actual class	Predicted background	Predicted wildfire occurrence
Background	373	12
Recorded wildfire occurrence	34	83

test set, the model correctly classified 373 of 385 background locations and 83 of 117 recorded wildfire occurrence locations, yielding an overall accuracy of 90.84%.

The performance metrics are summarised in Table 6. The model achieved an accuracy of 90.84%, precision of 87.37%, recall of 70.94%, specificity of 96.88%, F1-score of 78.30%, and ROC–AUC of 0.939. The high specificity indicates strong discrimination of sampled background locations, whereas the lower recall shows that a proportion of recorded wildfire occurrence locations were not correctly identified. Because the background locations do not represent confirmed fire absence, these metrics should be interpreted as discrimination between recorded wildfire occurrence locations and sampled background locations.

The ROC–AUC of 0.939 on the independent test set was comparable to the mean five-fold cross-validation ROC–AUC of 0.937 ± 0.014 obtained during model development within the training dataset. The strong performance is consistent with the ability of Random Forest to model complex and non-linear relationships among predictors without requiring parametric assumptions about their distributions (Breiman, 2001; Rodrigues and De la Riva, 2014). However, this agreement does not establish geographic generalisation. Random partitioning can place geographically proximate or environmentally similar observations in both training and testing subsets, potentially resulting in optimistic estimates of performance.

The random-split results were therefore considered benchmark performance estimates rather than evidence of spatially independent predictive capability. A spatial-block cross-validation procedure was subsequently used to provide a more stringent assessment of geographic generalisation.

Table 6. Performance metrics of the Random Forest model under stratified random hold-out validation

Metric	Value
Accuracy	90.84%
Precision	87.37%
Recall (Sensitivity)	70.94%
Specificity	96.88%
F1-score	78.30%
ROC–AUC	0.939
Mean five-fold CV ROC–AUC	0.937 ± 0.014

Spatial-block cross-validation of random forest performance

To assess geographic generalisation while reducing spatial dependence between training and validation observations, five-fold spatial-block cross-validation was performed using 10×10 km blocks. Blocks were assigned exclusively to individual folds using a stratified group-based procedure, preventing observations from the same block from occurring in both training and validation subsets (Figure 4; Table 7).

The spatial evaluation included 51 occupied blocks and 1,592 observations, comprising all 584 wildfire occurrence locations and 1,008 background locations. The remaining 918 background locations were excluded from this evaluation. Across the five folds, mean accuracy was $71.37 \pm 9.28\%$, precision $51.71 \pm 27.16\%$, recall $31.06 \pm 23.75\%$, specificity $90.52 \pm 7.33\%$, F1-score $36.43 \pm 21.75\%$, and ROC–AUC 0.714 ± 0.111 . Pooled out-of-fold predictions yielded an accuracy of 69.22%, precision 64.33%, recall 36.13%, specificity 88.39%, F1-score 46.27%, and ROC–AUC 0.706.

Compared with the random hold-out results (accuracy = 90.84%, F1-score = 78.30%, ROC–AUC = 0.939), spatial-block validation produced substantially lower performance, suggesting that the conventional random split benefited, at least partly, from spatial and environmental similarity between training and testing observations. Nevertheless, the pooled ROC–AUC of 0.706 indicates moderate discrimination under spatially separated validation.

Relative importance and explainability of predictors

Gini feature importance

The relative importance of the predictor variables derived from the Random Forest model is presented in Table 8 and Figure 5. Altitude had the highest Gini-based importance (0.158), followed by temperature (0.152) and NDVI (0.131), indicating that these variables contributed most strongly to the reduction of node impurity within the fitted forest. In Mediterranean mountain ecosystems, elevation controls temperature and precipitation gradients, which regulate vegetation distribution, fuel moisture, and fire susceptibility (Pausas and

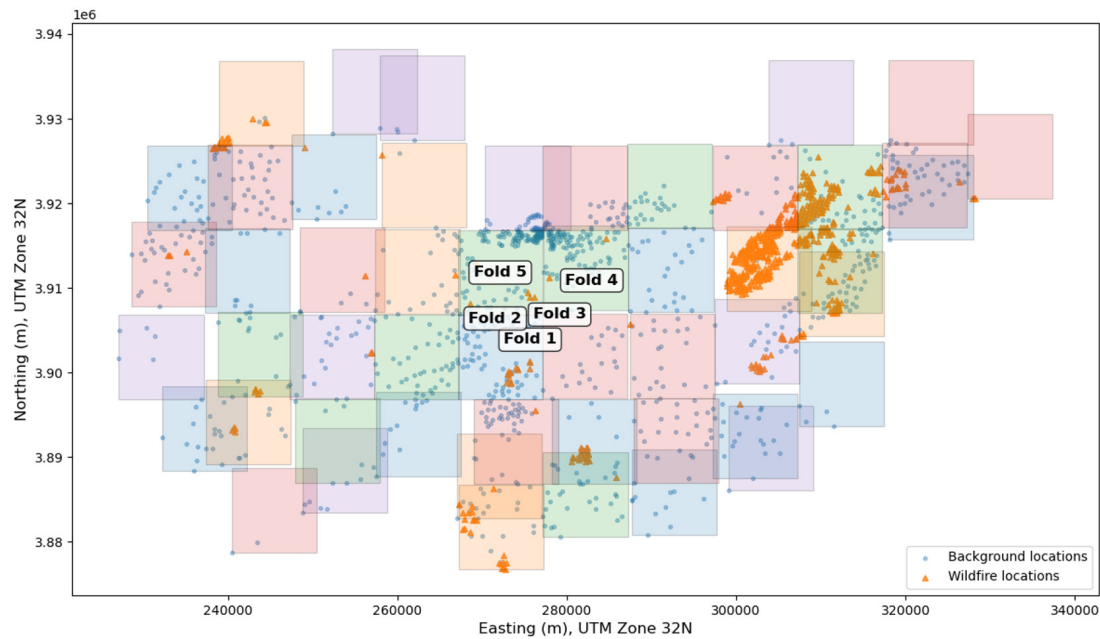


Figure 4. Spatial-block cross-validation scheme used to assess the geographic generalisation of the Random Forest model of recorded wildfire occurrence. The study observations were assigned to 10×10 km spatial blocks, with each block allocated exclusively to one of five validation folds. Orange triangles represent recorded wildfire occurrence locations and blue points represent background locations. Spatial blocking prevents observations within the same block from being simultaneously included in the training and validation datasets, thereby reducing potential local spatial dependence

Fernández-Muñoz, 2012; Ruffault et al., 2018). Anthropogenic variables also showed substantial importance, with distance to roads and distance to residential areas both receiving a Gini importance of 0.115. Rather than representing direct ignition causes, these variables can be interpreted as spatial proxies for human accessibility and associated activities, including transport, recreation, agriculture, and other human uses of the landscape (Costafreda-Aumedes et al., 2017; Kahli et al., 2026). Precipitation (0.115) and NDMI (0.081) showed moderate contributions, consistent with the role of moisture conditions in determining vegetation dryness and fuel flammability. In contrast, aspect

Table 8. Gini importance of the environmental and anthropogenic predictors used in the Random Forest model

Rank	Predictor	Gini importance
1	Altitude	0.158
2	Temperature	0.152
3	NDVI	0.131
4	Precipitation	0.115
5	Distance to roads	0.115
6	Distance to residential areas	0.115
7	NDMI	0.081
8	Aspect	0.053
9	Profile curvature	0.043
10	Slope	0.036

Table 7. Composition of the five spatial validation folds based on 10×10 km spatial blocks

Fold	Validation blocks	Recorded wildfire occurrence	Background	Total	Wildfire occurrence (%)
1	12	39	190	229	17.03
2	8	242	76	318	76.10
3	9	193	387	580	33.28
4	13	87	263	350	24.86
5	9	23	92	115	20.00
Total	51	584	1,008	1,592	36.68

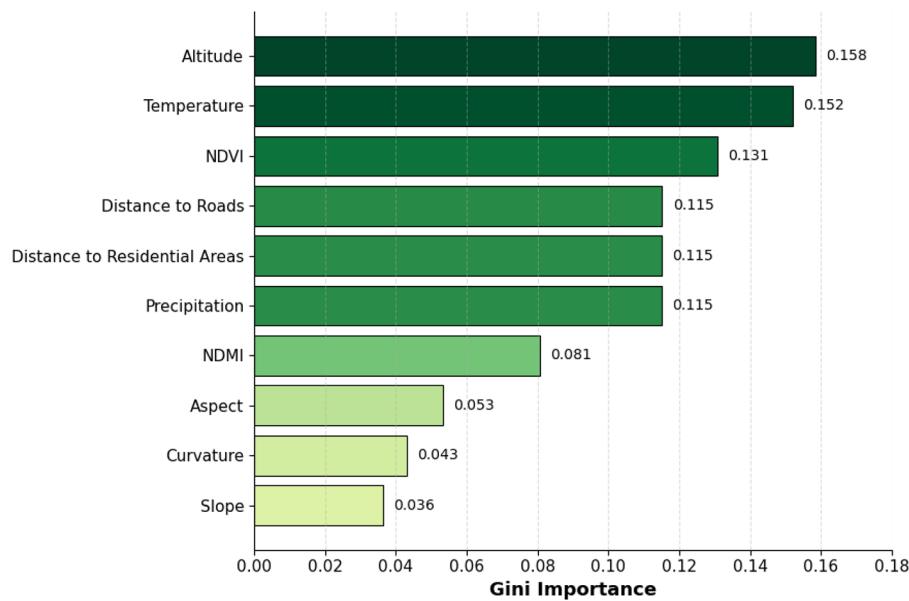


Figure 5. Relative Gini importance of the environmental and anthropogenic predictors in the Random Forest model of recorded wildfire occurrence. Higher values indicate greater contributions to impurity reduction across the fitted decision trees

(0.053), curvature (0.043), and slope (0.036) had relatively low Gini importance, indicating that these variables contributed less to the model's discrimination than the major climatic, vegetation, and accessibility-related predictors.

Permutation feature importance

The PFI analysis showed a broadly similar ranking of the leading predictors, although the relative importance of several secondary variables differed from the Gini-based results (Table 9; Figure 6). Altitude had the highest permutation importance (0.151 ± 0.015), followed by temperature (0.134 ± 0.011). Distance to residential areas

ranked third (0.056 ± 0.010), followed by precipitation (0.038 ± 0.008), distance to roads (0.038 ± 0.007), and NDVI (0.035 ± 0.008). NDMI showed a comparatively low importance (0.007 ± 0.005), while aspect, slope, and profile curvature contributed little to predictive performance. The low PFI of NDMI is consistent with its strong correlation with NDVI ($r = 0.868$), suggesting that the two predictors contain overlapping information and that NDVI may provide much of the predictive information shared by this vegetation-related predictor group. Thus, its low permutation importance reflects limited additional predictive information available to the fitted model rather than ecological unimportance. Differences between the

Table 9. Permutation feature importance of the environmental and anthropogenic predictors used in the Random Forest model

Rank	Predictor	Mean permutation importance	Standard deviation
1	Altitude	0.151	0.015
2	Temperature	0.134	0.011
3	Distance to residential areas	0.056	0.010
4	Precipitation	0.038	0.008
5	Distance to roads	0.038	0.007
6	NDVI	0.035	0.008
7	NDMI	0.007	0.005
8	Aspect	0.004	0.002
9	Slope	0.001	0.001
10	Profile curvature	−0.000	0.002

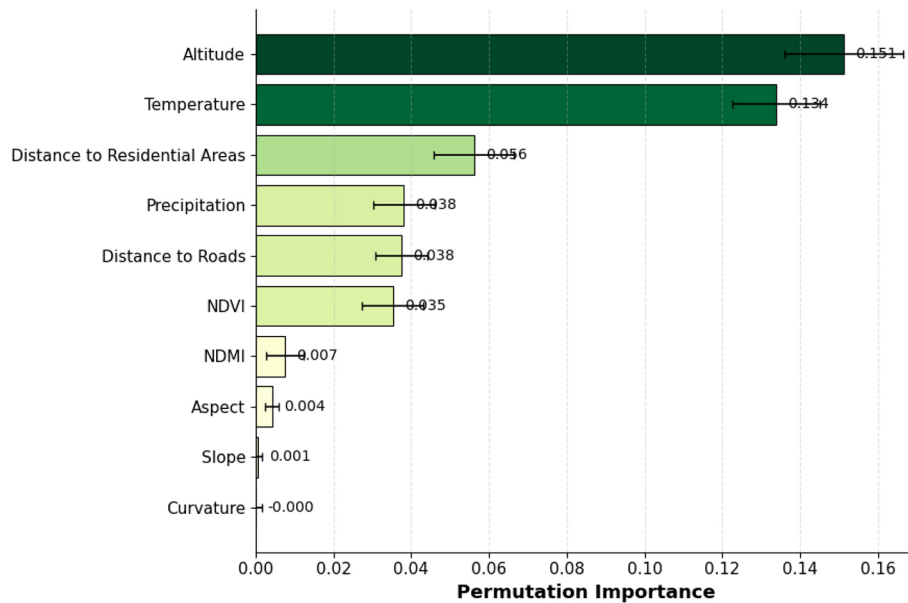


Figure 6. Permutation feature importance of the environmental and anthropogenic predictors in the Random Forest model. Bars represent mean decreases in predictive performance following permutation of each predictor, and error bars indicate standard deviations across 30 permutation repetitions

Gini and PFI rankings, particularly for distance to residential areas, distance to roads, and NDVI, are expected because the two measures quantify different aspects of model importance. Gini importance reflects impurity reduction during tree construction, whereas PFI measures the deterioration in predictive performance when a predictor is permuted while the remaining predictors are retained (Breiman, 2001; Fisher et al., 2019).

SHAP-based interpretation of predictor contributions

SHAP analysis provided an additional assessment of the magnitude and direction of individual predictor contributions to the modelled recorded-wildfire occurrence score. Based on mean absolute SHAP values, altitude (0.090) and temperature (0.080) had the largest overall contributions

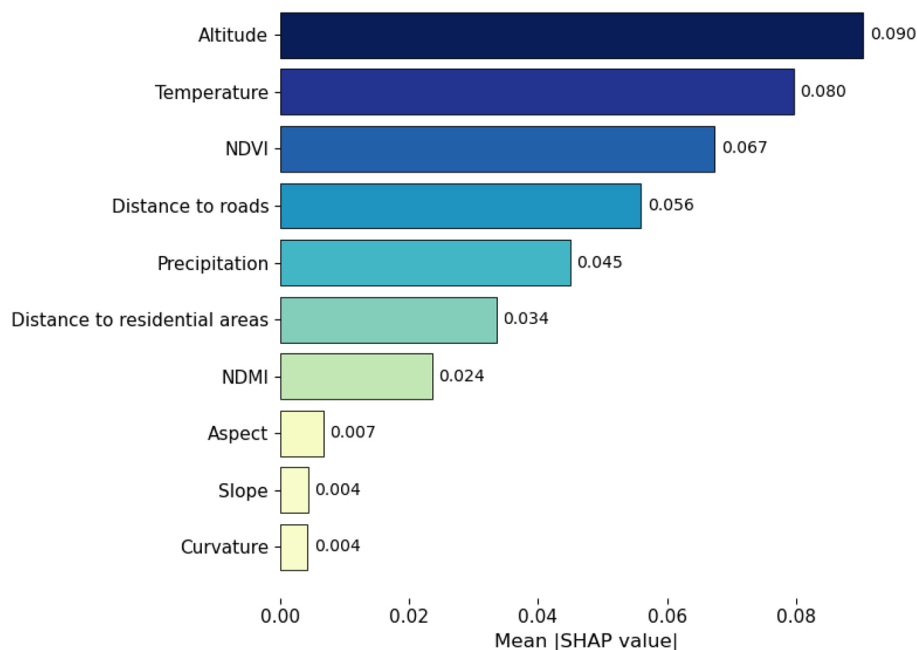


Figure 7. Global SHAP feature importance for the Random Forest model of recorded wildfire occurrence. Mean absolute SHAP values represent the average magnitude of each predictor's contribution to model predictions

to model predictions, followed by NDVI (0.067), distance to roads (0.056), precipitation (0.045), and distance to residential areas (0.034) (Figure 7). NDMI showed a smaller contribution (0.024), while aspect, slope, and profile curvature had comparatively limited influence. SHAP values further indicated that higher temperature generally increased the modelled wildfire-occurrence score, whereas higher precipitation and NDMI generally reduced it. Shorter distances to roads and residential areas were also associated with higher modelled scores. These patterns reflect the predictive behaviour of the fitted RF model rather than causal relationships. In particular, the contribution attributed to altitude should be interpreted alongside its strong association with temperature and precipitation, while the contributions of NDVI and NDMI may partly reflect their shared information.

CONCLUSION

This study developed an explainable Random Forest framework to investigate the spatial distribution of recorded wildfire occurrence in the Aurès Mountains of north-eastern Algeria using multi-source geospatial data. The model showed strong discrimination under conventional stratified random partitioning, achieving an accuracy of 90.84% and an ROC–AUC of 0.939. However, performance declined substantially under spatial-block cross-validation, with a mean ROC–AUC of 0.714 ± 0.111 and a pooled out-of-fold ROC–AUC of 0.706, indicating that conventional random partitioning may provide optimistic estimates of predictive performance when spatial dependence is present.

The XAI analysis consistently identified altitude and temperature among the most important predictors, while vegetation, precipitation, and accessibility-related variables also contributed to model discrimination. SHAP analysis indicated that higher temperature and NDVI values were generally associated with higher modelled recorded-wildfire occurrence scores, whereas higher precipitation and NDMI values were generally associated with lower scores. Shorter distances to roads and residential areas also contributed to higher modelled scores. These patterns represent model-dependent spatial associations rather than independent or causal effects,

particularly because several predictors shared substantial information.

The findings should be interpreted in the context of the response data, which represent satellite-recorded active-fire detections rather than a complete inventory of all wildfire events. Similarly, background locations represent sampled locations without a recorded FIRMS detection and should not be interpreted as confirmed fire absences. The framework therefore characterises spatial discrimination between recorded wildfire occurrence locations and sampled background conditions within the environmental domain represented by the study area and dataset.

Despite these limitations, the results provide a useful basis for identifying areas that may warrant increased wildfire monitoring, prevention, and management attention in the Aurès Mountains. Future studies should incorporate temporally matched weather and drought indicators, fuel-moisture conditions, and more detailed measures of human activity, together with independent geographic validation, to improve the temporal relevance and spatial transferability of wildfire susceptibility assessments across Mediterranean mountain environments.

REFERENCES

1. Ballais, J.-L., Cohen, J. (1981). Intérêt morphogénétique et paléoclimatique des travertins des Aurès (Algérie). In *Formations carbonatées externes, tufs et travertins* (pp. 37–44). Association des Géographes Français.
2. Beghami, Y., Kalla, M., Thion, M., Benmessaoud, H. (2012). Spatiotemporal dynamics of forest and mountain formations in Aurès area, Algeria. *Journal of Life Sciences*, 6(6), 663–669.
3. Belgiu, M., Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
4. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
5. Chuvieco, E., Aguado, I., Salas, J., García, M., Yebra, M., Oliva, P. (2020). Satellite remote sensing contributions to wildland fire science and management. *Current Forestry Reports*, 6(2), 81–96.
6. Chuvieco, E., Mouillot, F., van der Werf, G. R., San Miguel, J., Tanasse, M., Koutsias, N., García, M., Yebra, M., Padilla, M., Gitas, I., Heil, A., Hawbaker,

- T. J., Giglio, L. (2019). Historical background and current developments for mapping burned area from satellite Earth observation. *Remote Sensing of Environment*, 225, 45–64. <https://doi.org/10.1016/j.rse.2019.02.013>
7. Costafreda-Aumedes, S., Comas, C., Vega-García, C. (2017). Human-caused fire occurrence modelling in perspective: A review. *International Journal of Wildland Fire*, 26(12), 983–998. <https://doi.org/10.1071/WF17026>
 8. Curt, T., Aini, A., Dupire, S. (2020). Fire activity in Mediterranean forests (The Algerian case). *Fire*, 3(4), 58. <https://doi.org/10.3390/fire3040058>
 9. Curt, T., Fréjaville, T., Lahaye, S. (2016). Modelling the spatial patterns of ignition causes and fire regime features in southern France: Implications for fire prevention policy. *International Journal of Wildland Fire*, 25(7), 785–796.
 10. Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
 11. Fick, S. E., Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
 12. Fisher, A., Rudin, C., Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177), 1–81.
 13. Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
 14. Intergovernmental Panel on Climate Change (IPCC). (2023). *Climate change 2023: Synthesis report*. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (H. Lee & J. Romero, Eds.). IPCC. <https://doi.org/10.59327/IPCC/AR6-9789291691647>
 15. James, G., Witten, D., Hastie, T., Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
 16. Kahli, A., Sahnouni, R., Saoula, H., Chenchouni, H. (2026). Modelling environmental drivers and socioeconomic/human predictors of forest fire outbreaks in Algeria. *Natural Hazards*, 122(8), 334. <https://doi.org/10.1007/s11069-026-08061-5>
 17. Lundberg, S. M., Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc.
 18. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>
 19. Mallinis, G., Verde, N., Siachalou, S., Latinopoulos, D., Akratos, C., Kagalou, I. (2023). Evaluation of multiple classifier systems for mapping different hierarchical levels of forest ecosystems in the Mediterranean region using Sentinel-2, Sentinel-1, and ICESat-2 data. *Forests*, 14(11), 2224. <https://doi.org/10.3390/f14112224>
 20. Molnar, C. (2022). *Interpretable machine learning* (2nd ed.). Lulu.com. <https://christophm.github.io/interpretable-ml-book/>
 21. Moreira, F., Ascoli, D., Safford, H., Adams, M. A., Moreno, J. M., Pereira, J. M. C., Catry, F. X., Armesto, J., Bond, W., González, M. E., Curt, T., Koutsias, N., McCaw, L., Price, O., Pausas, J. G., Rigolot, E., Stephens, S., Tavşanoğlu, Ç., Vallejo, R., Fernandes, P. M. (2020). Wildfire management in Mediterranean-type regions: Paradigm change needed. *Environmental Research Letters*, 15(1), 011001. <https://doi.org/10.1088/1748-9326/ab541e>
 22. Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., Pereira, J. M. C. (2012). Modelling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. *Forest Ecology and Management*, 275, 117–129. <https://doi.org/10.1016/j.foreco.2012.03.003>
 23. Pausas, J. G., Fernández-Muñoz, S. (2012). Fire regime changes in the Western Mediterranean Basin: From fuel-limited to drought-driven fire regime. *Climatic Change*, 110(1), 215–226. <https://doi.org/10.1007/s10584-011-0060-6>
 24. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
 25. Rodrigues, M., De la Riva, J. (2014). An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software*, 57, 192–201. <https://doi.org/10.1016/j.envsoft.2014.03.003>

26. Rossa, C. G., Veloso, R., Fernandes, P. M. (2015). A laboratory-based quantification of the effect of slope on fire spread rate. *International Journal of Wildland Fire*, 24(2), 191–200. <https://doi.org/10.1071/WF14035>
27. Rouse, W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium* (Vol. 1, pp. 309–317). NASA.
28. Ruffault, J., Curt, T., Martin-StPaul, N. K., Moron, V., Trigo, R. M. (2018). Extreme wildfire events are linked to global-change-type droughts in the northern Mediterranean. *Natural Hazards and Earth System Sciences*, 18(3), 847–856. <https://doi.org/10.5194/nhess-18-847-2018>
29. Ruffault, J., Curt, T., Moron, V., Trigo, R. M., Mouillot, F., Koutsias, N., Pimont, F., Martin-StPaul, N. K., Dupuy, J.-L. (2020). Increased likelihood of heat-induced large wildfires in the Mediterranean Basin. *Scientific Reports*, 10, 13790. <https://doi.org/10.1038/s41598-020-70069-z>
30. San-Miguel-Ayanz, J., Durrant, T., Boca, R., Mairanti, P., Libertà, G., Oom, D., Branco, A., De Rigo, D., Suarez-Moreno, M., Ferrari, D., Roglia, E., Scionti, N., Broglia, M., Onida, M., Tistan, A., Löffler, P. (2024). *Forest fires in Europe, Middle East and North Africa 2023*. Publications Office of the European Union. <https://doi.org/10.2760/8027062>
31. Trucchia, A., D’Andrea, M., Fiorucci, P., Negro, D., Vitolo, C. (2022). Machine learning applications for wildfire susceptibility modelling: A review. *Environmental Modelling & Software*, 153, 105443.
32. Turco, M., Rosa-Cánovas, J. J., Bedia, J., Jerez, S., Montávez, J. P., Llasat, M. C., Provenzale, A. (2018). Exacerbated fires in Mediterranean Europe due to anthropogenic warming projected with non-stationary climate-fire models. *Nature Communications*, 9, 3821. <https://doi.org/10.1038/s41467-018-06358-z>
33. Valavi, R., Elith, J., Lahoz-Monfort, J. J., Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. *Ecography*, 44(12), 1731–1742. <https://doi.org/10.1111/ecog.05615>