EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology, 2026, 27(6), 259–273
https://doi.org/10.12912/27197050/221629
ISSN 2719–7050, License CC-BY 4.0

Received: 2026.04.13
Accepted: 2026.05.15
Published: 2026.06.01

Machine learning-based wildfire occurrence prediction using integrated meteorological and fire weather indicators: A case study of northern Morocco

Chaimae Ouazri[1*], Abderrahim El Mhouti[1], Mohamed Fahim[1], Nisrine El Ayat[2]

[1] ISISA Lab, Faculty of Science, Abdelmalek Essaadi University, Tetouan, Morocco
[2] ELITT-Lab, Higher School of Technology, Abdelmalek Essaadi University, Tetouan, Morocco
* Corresponding author's e-mail: couazri@uae.ac.ma

ABSTRACT

Wildfire occurrence has intensified across Mediterranean ecosystems under the combined effect of rising temperatures and prolonged summer droughts, posing a critical challenge for fire risk management in data-scarce regions such as North Africa. This study proposes a machine learning framework for predicting wildfire occurrence in northern Morocco using integrated meteorological and fire weather indicators. A multi-source dataset covering the Tangier–Tétouan–Al Hoceima region was constructed by combining ERA5 meteorological reanalysis variables and fire weather index (FWI) components with wildfire occurrence data derived from NASA FIRMS for the period 2019–2024. Two gradient boosting algorithms, XGBoost and LightGBM, were trained using a temporal hold-out strategy (train: 2019–2022, test: 2023–2024), and a structured ablation study was conducted to quantify the contribution of each feature group. Results show that XGBoost trained on the full feature set achieved the best performance (accuracy = 0.737, ROC-AUC = 0.813, PR-AUC = 0.739). The ablation analysis demonstrates that FWI components significantly improve predictive performance by capturing multi-day fuel moisture dynamics not represented by instantaneous meteorological conditions, while meteorological and FWI variables alone retain meaningful predictive skill in the absence of geographic coordinates (ROC-AUC = 0.733). Post-hoc SHAP analysis identifies longitude, surface pressure, temperature, and solar radiation as the dominant predictors, with longitude reflecting spatially structured fire occurrence patterns consistent with multiple co-varying geographic factors across the TTA region. The findings provide an interpretable and temporally validated machine learning framework that can support early warning and fire risk management in Mediterranean North Africa.

Keywords: wildfire prediction, XGBoost, fire weather index, ERA5, northern Morocco, SHAP, gradient boosting.

INTRODUCTION

Background

Wildfires constitute one of the most destructive natural hazards affecting Mediterranean-climate ecosystems, causing severe ecological damage, biodiversity loss, carbon emissions, and significant socioeconomic consequences. The Mediterranean Basin spanning southern Europe, North Africa, and the Middle East is recognized as a global fire-prone hotspot, where the convergence of hot, dry summers and fuel-rich vegetation creates conditions highly conducive to large-scale fire events (Jain et al., 2020). Within this broader geographic context, Morocco in particular has experienced a marked intensification of wildfire activity over recent decades. According to EFFIS, Morocco recorded over 500 wildfires that burned approximately 22,700 hectares in 2022 alone, with the mountainous Rif region consistently ranking among the most severely affected areas. The Talassemtane National Park hosts rare endemic species including the Moroccan fir (*Abies marocana*) and the Barbary macaque (*Macaca sylvanus*), both critically threatened by recurring fire disturbances (Castro et al., 2022). Climate change is fundamentally

altering the fire regime in this region and globally. Rising temperatures, prolonged droughts, and increasingly frequent heatwaves are intensifying conditions that favor wildfire ignition and spread. Abatzoglou et al. (2025) demonstrated that human-caused climate change has increased the frequency of extreme fire-weather years by 88–152% across global forested lands compared to preindustrial conditions. Warming trends are projected to significantly lengthen fire seasons and expand high-risk areas, placing underrepresented sub-regions such as North Africa at the forefront of climate-driven fire risk escalation (Jones et al., 2022).

Existing approaches

Traditional wildfire risk assessment has historically relied on physical fire danger indices, most notably the Canadian forest fire weather index (FWI) system a meteorologically-based composite index integrating temperature, relative humidity, wind speed, and precipitation through a hierarchical system of moisture codes and behavioral indices (Van Wagner, 1987). The availability of FWI computed from ERA5 reanalysis data produced by ECMWF has enabled gridded, temporally consistent fire danger datasets at global scale (Di Giuseppe et al., 2020).

Despite its widespread deployment, the FWI system operates on simplified physical assumptions: it is calibrated for uniform boreal vegetation and neglects topographic heterogeneity, land-cover variability and the stochastic nature of ignition events. ERA5-derived FWI values systematically underestimate peak fire danger relative to ground-based observations, particularly in regions with sparse weather station networks (Sjöström et al., 2025). These limitations become especially pronounced in complex terrain such as the Rif Mountains, where elevation gradients, fuel type diversity and Mediterranean micro-climates interact in ways that index-based approaches fail to represent.

In response to these shortcomings, machine learning (ML) methods have emerged as a powerful paradigm for wildfire prediction, offering the capacity to model nonlinear, high-dimensional relationships between environmental drivers and fire occurrence. Ensemble tree models particularly Random Forest (RF) and XGBoost have shown superior accuracy over classical statistical methods in fire occurrence classification tasks (Iban and Sekertekin, 2022), with AUC values routinely exceeding 0.90 (Moghim and Mehrabi, 2024). Concurrently, SHapley Additive exPlanations (SHAP) have enabled practitioners to decompose model predictions into per-feature contributions, addressing the "black-box" opacity that limits operational trust (Lundberg and Lee, 2017). SHAP-based interpretability has been successfully applied in Mediterranean wildfire contexts, consistently identifying meteorological variables as the most influential predictors (Cilli et al., 2022).

In the TTA region specifically, Moumane et al. (2025) applied XGBoost and LightGBM with SHAP to map wildfire susceptibility, identifying NDVI and wind speed as dominant drivers. Mohajane et al. (2021) developed hybrid ML models for forest fire susceptibility mapping in Northern Morocco, achieving AUC values up to 0.989 using topographic, climatic and vegetation variables.

Research gap

Despite significant advances in ML-based wildfire modeling at the global and northern Mediterranean scales, North Africa remains critically underrepresented in the scientific literature. A recent bibliometric review of 138 wildfire risk prediction studies identified only a single study focused on Morocco and one on Algeria (Xu et al., 2025). The study of Zaidi (2023) explicitly acknowledged the lack of publicly available datasets for the region, underscoring the need for studies leveraging globally available gridded datasets such as ERA5 to bridge this data gap.

From a methodological standpoint, the joint integration of ERA5-derived meteorological variables with FWI sub-components within a unified ML framework has received limited attention in the North African context. Most existing studies either rely solely on fire weather indices or on meteorological variables, preventing a systematic evaluation of each input group's contribution. Furthermore, few studies have conducted structured ablation analyses to quantify the marginal predictive utility of individual feature groups, leaving open the question of which predictor categories are truly informative in this specific fire regime.

To address these methodological and geographic gaps, this study proposes an ML-based framework for daily wildfire occurrence prediction in northern Morocco, with four principal contributions. First, a unified gridded dataset

(0.25°, 2019–2024) was constructed by combining ERA5 reanalysis variables, FWI components and MODIS FIRMS fire occurrence labels the first open multi-source wildfire dataset for the Tangier-Tétouan-Al Hoceima region. Second, a systematic feature ablation study quantifies the marginal predictive value of FWI components versus raw meteorological variables across four configurations (Full, No_FWI, No_Meteo, and No_Spatial) evaluated on both XGBoost and LightGBM. Third, baseline performance metrics (PR-AUC = 0.739, ROC-AUC = 0.813) are established for daily fire occurrence prediction in the TTA region. Fourth, SHAP analysis identifies and ranks the dominant fire drivers longitude, surface pressure, temperature, solar radiation and drought code (DC) providing transparent feature attribution for operational fire danger forecasting. The remainder of this paper is organized as follows: Section 2 describes the study area; Section 3 details data and feature engineering; Section 4 presents the ML models and experimental design; Section 5 reports results and SHAP analysis; Sections 6 and 7 provide discussion and conclusions.

METHODOLOGY

Study area

The Tangier-Tétouan-Al Hoceima (TTA) region is situated in the northernmost part of Morocco (longitudes 6°14′36″W–3°48′12″W, latitudes 34°30′34″N–35°55′21″N), covering approximately 17,262 km² with dual maritime exposure to the Mediterranean Sea and the Atlantic Ocean (Boubekraoui et al., 2023). The region is dominated by the Rif mountain chain (maximum elevation 2.456 m at Jbel Tidighine), producing pronounced altitudinal and microclimatic gradients that strongly influence vegetation composition and wildfire dynamics (Figure 1). With a population exceeding 4 million inhabitants (≈233 inhabitants/km²), ongoing urban expansion intensifies anthropogenic pressure on natural ecosystems, increasing the frequency of human-caused fire ignitions (Boubekraoui et al., 2023).

The TTA region experiences a classic Mediterranean climate (Köppen Csa/Csb), with June–September constituting the core fire weather season marked by atmospheric drying, elevated temperatures, and low relative humidity (Salhi et al., 2019; Driouech et al., 2021). Rainfall is highly heterogeneous and modulated by orography, ranging from approximately 750 mm in Tetouan to only 350–400 mm near Al Hoceima. Long-term observations indicate a strengthening of summer aridity consistent with broader projections for the southern Mediterranean anticipating a significant lengthening of the fire weather season by mid-century.

Vegetation is organized along three altitudinal bands: the submontane zone (below 600 m) dominated by cork oak (*Quercus suber*) and matorral forming highly flammable fuel beds; the montane

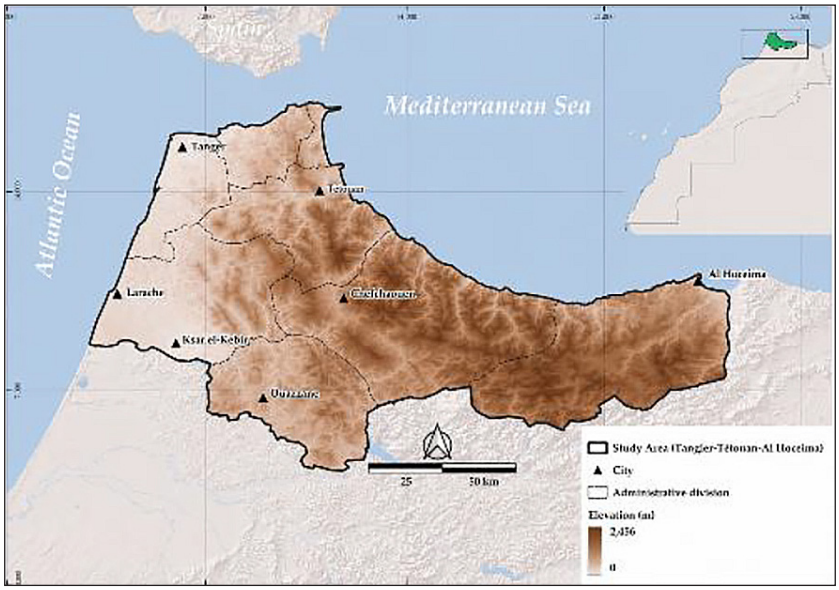

Figure 1. Study area located in the Tangier–Tetouan–Al Hoceima (TTA) administrative region, northern Morocco

zone (600–1,400 m) characterized by mixed oak and pine stands; and the subalpine zone (above 1.400 m) hosting relic populations of Atlas cedar (*Cedrus atlantica*) and Moroccan fir (*Abies marocana*) within Talassemtane National Park, forming part of the Betic-Rifean biodiversity hotspot (Maranón et al., 2004). Two major protected areas Talassemtane National Park (596 km²) and Al Hoceima National Park (282 km²) have been repeatedly affected by fire events (Castro et al., 2022).

The TTA region consistently ranks first nationally in wildfire frequency and burned area (Boubekraoui et al., 2023). Analysis of MODIS Fire_CCI51 (2002–2020) and FIRMS data (2001–2022) reveals an average of 39.78 km²/year of burned area, concentrated in forests (74%) and matorral (Figure 2), with the Rif sub-region accounting for approximately 45% of national fire departures.

The summer of 2022 was particularly catastrophic, with approximately 500 wildfires burning over 22,700 hectares nationally, underscoring the urgent need for a data-driven fire prediction system calibrated for this region.

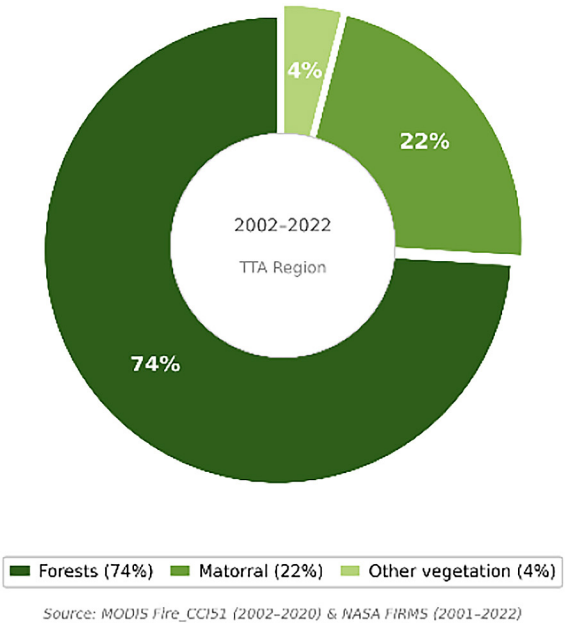

Figure 2. Burned area by vegetation type in the Tangier–Tétouan–Al Hoceima (TTA) region, 2002–2022

Data sources

ERA5 meteorological reanalysis

ERA5 is the fifth-generation global atmospheric reanalysis produced by ECMWF for the Copernicus Climate Change Service (Hersbach et al., 2020), providing hourly estimates of atmospheric variables at 0.25° spatial resolution (~31 km) from 1940 to near-present. Daily aggregates were extracted for the TTA region (longitude: 6.25°W–3.75°W; latitude: 34.5°N–36.0°N) over 2019–2024 from the Copernicus Climate Data Store. Seven variables were retrieved for the TTA region: 2 m temperature (t2m) and dewpoint temperature (d2m), used to derive relative humidity; U and V wind components at 10 m (u10, v10), used to derive scalar wind speed, total precipitation (tp), surface pressure (sp) and surface net solar radiation (ssr). Full descriptions are provided in Table 1 (Group B).

Fire weather index components

The Canadian forest fire weather index (FWI) system provides a hierarchical set of fuel moisture codes and fire behavior indices computed from daily meteorological inputs (Van Wagner, 1987). Daily FWI outputs were obtained from the Global ECMWF Fire Forecast dataset at 0.25° resolution via the Copernicus Climate Data Store, computed using ERA5 inputs (Di Giuseppe et al., 2020) to ensure spatio-temporal consistency with the meteorological predictors. Six FWI components were used: three moisture codes (FFMC, DMC, DC) tracking fuel dryness at different depths and timescales from surface litter (FFMC) to moderately deep organic layer (DMC) to long-term drought conditions (DC) two fire behavior indices (ISI, estimating initial spread rate, and BUI, representing total available fuel) and the composite FWI danger rating. Full descriptions of all components are provided in Table 1 (Group C)

Fire occurrence labels

Binary fire occurrence labels (fire = 1, no-fire = 0) were derived from NASA's Fire Information for Resource Management System (FIRMS), using MODIS active fire detections from Terra and Aqua satellites (MCD14ML Collection 6.1, 1 km resolution). Detections over the TTA region (2019–2024) were spatially matched to the 0.25° ERA5-FWI grid. A grid cell-day was labeled as fire if at least one detection with confidence ≥ 50% occurred within the cell on that day (Jain et al., 2020). The dataset was subsequently balanced to achieve equal representation of fire and non-fire events (7.732 instances per class).

Feature construction

Relative humidity

ERA5 provides 2 m temperature (T_{2m}, in Kelvin) and 2 m dewpoint temperature (Td_{2m}) as standard output fields, but not relative humidity (RH) directly. Relative humidity was computed from these fields using the Magnus-Tetens approximation. Temperature values were first converted from Kelvin to Celsius. The saturation vapor pressure ratio was then computed as:

$$RH = 100.exp\left(\frac{17.625.Td}{243.04+Td} - \frac{17.625.T}{243.04+T}\right) \quad (1)$$

where: T and Td are temperature and dewpoint in °C.

This formulation provides accurate estimates of relative humidity for the temperature range relevant to surface fire weather conditions and is widely used in meteorological applications.

Wind speed

ERA5 provides the horizontal wind field at 10 m above ground level as two Cartesian components: the U-component (u_{10}, positive eastward) and the V-component (v_{10}, positive northward). Scalar wind speed was derived as:

$$WS = \sqrt{(u_{10}{}^2 + v_{10}{}^2)} \quad (2)$$

Wind direction was not included in the prediction feature set, as its relationship to fire ignition probability is non-monotonic and direction-dependent in the complex Rif terrain.

Final feature groups

The final dataset consists of 15 predictors organized into three semantically coherent feature groups spatial (Group A), meteorological (Group B) and fire weather indices (Group C) plus the binary response variable (fire/no-fire). This three-group structure was intentionally designed to support the structured ablation analysis, wherein each group is selectively excluded to quantify its marginal contribution to predictive performance. Table 1 provides a complete description of all features, including their units and roles.

Several design decisions embedded in this feature set merit explicit justification. First, the simultaneous inclusion of both raw meteorological variables (Group B) and FWI components (Group C) is deliberate: although the FWI sub-components are themselves functions of meteorological inputs, they encode temporally integrated moisture state particularly through the DC and DMC, which track multi-day and multi-week drying that is not captured by the instantaneous daily weather snapshot. The two groups are therefore complementary rather than redundant and their joint contribution is precisely what the ablation study aims to quantify. Second, NDVI and topographic variables (elevation, slope, aspect) were intentionally excluded from the primary feature set to maintain focus on meteorological predictability and to avoid conflating fire weather prediction with fire susceptibility mapping. Third, the spatial resolution of all features (0.25°) was kept uniform across Groups A, B and C to avoid resolution-induced artifacts in the training data.

Experimental design

Train/test split

The dataset was balanced through random undersampling prior to temporal splitting to ensure equal representation of fire and non-fire events (7,732 instances per class). Consequently, both the training (2019–2022) and test (2023–2024) sets follow a balanced class distribution. To prevent data leakage arising from the temporal autocorrelation inherent in daily meteorological observations particularly the multi-week memory encoded in FWI moisture codes such as the DC and duff moisture code (DMC) a temporal hold-out strategy was applied. While this balanced setup does not reflect real-world fire prevalence, where fire events represent a small fraction of observations (~2–5%), it enables consistent comparison across all ablation configurations. To address this limitation, an additional evaluation under the natural class distribution is presented in the Discussion (Operational Implications) section. Bootstrap confidence intervals (n = 1.000) are reported for primary metrics to assess result stability.

Consistency across ablation experiments

A critical methodological requirement is the strict comparability of all experimental conditions across models and feature configurations. To ensure this, the identical temporal partition (2019–2022 for training, 2023–2024 for testing) is reused across all eight experimental configurations (four feature sets × two models). No re-splitting, re-stratification, or data augmentation is performed

Table 1. Overview of the fifteen final predictors incorporated into the machine learning framework for wildfire occurrence prediction, categorized into three distinct feature groups

Group	Category	Feature name	Unit/Range	Role/Description
A	Spatial	Latitude (lat)	°N	Encodes geographic position and captures latitudinal gradients in wildfire risk.
A	Spatial	Longitude (lon)	°E	Encodes east–west spatial gradients influencing regional fire regimes in the TTA region.
B	Meteorological	2 m temperature (t2m)	°C	Primary driver of fuel drying processes and ignition thresholds.
B	Meteorological	Relative humidity (RH)	%	Derived from t2m and d2m; controls the moisture content of surface fuels.
B	Meteorological	Wind speed (wind_speed)	m s^{-1}	Derived from the u10 and v10 wind components; influences fire spread and oxygen supply.
B	Meteorological	Total precipitation (tp)	mm day^{-1}	Daily accumulated precipitation contributing to fuel moisture and reducing ignition probability.
B	Meteorological	Surface pressure (sp)	Pa	Atmospheric pressure influencing local meteorological conditions and drying processes.
B	Meteorological	Surface net solar radiation (ssr)	J m^{-2}	Net solar radiation reaching the surface; contributes to fuel drying and evapotranspiration.
C	Fire weather indices	Fine fuel moisture code (ffmcode)	0–101	Surface litter moisture indicator strongly associated with ignition probability.
C	Fire weather indices	Duff moisture code (dufmcode)	0–500+	Represents moisture in the moderately deep organic layer and medium-term drying conditions.
C	Fire weather indices	Drought code (drtcode)	0–1000+	Long-term drought indicator representing deep organic fuel dryness.
C	Fire weather indices	Initial spread index (infsinx)	0–50+	Estimates the expected rate of fire spread immediately after ignition.
C	Fire weather indices	Build-up index (fbupinx)	0–1000+	Represents the total amount of fuel available for combustion.
C	Fire weather indices	Fire weather index (fwinx)	0–100+	Composite fire danger indicator estimating potential fire intensity.
C	Fire weather indices	Daily severity rating (fdsrte/ DSR)	0–100+	This indicator quantifies daily fire severity and potential control difficulty, emphasizing extreme fire conditions.

between runs. Feature subsetting for the ablation study is applied post-split by column selection on the already-partitioned X_{train} and X_{test} matrices. An additional No_spatial configuration excluding latitude and longitude was evaluated to disentangle spatial memorization from genuine meteorological predictive skill. This design ensures that all metric differences are attributable solely to feature composition or model architecture and not to sampling variation in the data partition.

Machine learning models

XGBoost

The eXtreme Gradient Boosting classifier (XGBoost) is an optimized implementation of gradient-boosted decision trees that incorporates second-order Taylor expansion of the binary cross-entropy loss, explicit L1 and L2 regularization, and stochastic column and row subsampling to control variance. At each boosting round m, a new regression tree h_m is fitted to the negative gradient of the loss and added to the ensemble according to the update rule:

$$\hat{F}_m(x) = \hat{F}_{m-1}(x) + \eta \cdot h_m(x) \quad (3)$$

where: $\eta = 0.05$ is the learning rate.

The model is trained for 500 boosting iterations (n_estimators = 500) with a maximum tree depth of 6 (max_depth = 6). Binary cross-entropy is specified as the training loss via eval_metric = 'logloss'. Predicted probabilities are obtained using the predict_proba method and are used to compute threshold-independent evaluation metrics such as ROC-AUC and PR-AUC.

LightGBM

LightGBM is a gradient boosting framework that employs a leaf-wise (best-first) tree growth strategy splitting the single leaf with the maximum loss reduction at each step rather than the depth-wise (level-wise) approach used by XGBoost.

This strategy results in asymmetric trees that can converge faster and achieve lower training loss for a given model complexity.

In the present configuration, tree depth is explicitly bound to 6 levels via max_depth = 6 to maintain comparability with the XGBoost configuration, while the learning rate and number of estimators are matched at learning_rate = 0.05 and n_estimators = 500, respectively.

All remaining LightGBM parameters, including num_leaves, subsampling fractions, and regularization terms, are kept at their default values in the scikit-learn API.

As with XGBoost, predicted probabilities are obtained using predict_proba and are used for threshold-independent metric computation.

Hyperparameter configuration

For reproducibility purposes, the hyperparameters used for both models are reported in Table 2. Parameters not explicitly specified are set to their default values in the XGBoost and LightGBM implementations. A fixed random_state of 42 is applied to both models to ensure reproducibility of stochastic processes, including subsampling and tie breaking.

No hyperparameter optimization, such as grid search or randomized search, was performed, as the primary objective of this study is to evaluate the relative contribution of different feature groups within a controlled ablation framework rather than to maximize predictive performance.

Maintaining fixed hyperparameters ensures that performance differences can be attributed to feature composition rather than optimization variability, thereby improving the interpretability of the results.

Ablation study design

To quantify the marginal contribution of each feature group to predictive performance, a structured ablation study is conducted by training and evaluating both models on four distinct feature sets. The four configurations are defined by selective exclusion of feature groups from the full 15-feature set (Table 3).

The full configuration serves as the reference model. The No_FWI variant isolates the contribution of the FWI system by excluding all FWI components, thereby assessing their added value beyond instantaneous meteorological conditions. Conversely, the No_meteo configuration evaluates whether FWI components alone can capture the predictive information contained in raw meteorological variables, given that they are derived from the same underlying inputs. The No_spatial configuration additionally excludes geographic coordinates to quantify the contribution of spatial structure and assess whether meteorological and FWI variables retain sufficient predictive skill in the absence of location information.

Together, these configurations form a controlled ablation framework over the two non-spatial feature groups, enabling a systematic

Table 2. Explicitly configured hyperparameters for XGBoost and LightGBM as implemented in the pipeline. All unlisted parameters use library defaults

Model	Parameter	Value	Description
XGBoost	n_estimators	500	Number of boosting iterations used during training
	learning_rate	0.05	Learning rate controlling the contribution of each tree (η)
	max_depth	6	Maximum depth of individual trees
	subsample	0.8	Row subsampling ratio to reduce variance
	colsample_bytree	0.8	Column subsampling ratio per tree
	eval_metric	logloss	Binary cross-entropy loss used during training
	random_state	42	Fixed seed ensuring reproducibility
LightGBM	n_estimators	500	Number of boosting iterations
	learning_rate	0.05	Learning rate controlling tree contribution
	max_depth	6	Maximum depth constraint
	num_leaves	50	Maximum number of leaves per tree
	subsample	0.8	Row subsampling ratio
	colsample_bytree	0.8	Column subsampling ratio
	random_state	42	Fixed seed ensuring reproducibility

Table 3. Feature set configurations used in the ablation study.

Feature Set	Features Included	Purpose of the Ablation
Full	Spatial + meteorological + FWI	Uses all 15 predictors: latitude, longitude, t2m, RH, wind_speed, precipitation, pressure, solar radiation, and all FWI components (FFMC, DMC, DC, ISI, BUI, FWI, DSR). Serves as the reference model.
No_FWI	Spatial + meteorological	Excludes the all FWI components. Evaluates the marginal contribution of the FWI system beyond the raw meteorological variables.
No_meteo	Spatial + FWI	Excludes the direct meteorological predictors (t2m, RH, wind_speed, precipitation, surface pressure, and solar radiation). Tests whether FWI components alone implicitly encoding meteorological conditions are sufficient for wildfire prediction.
No_spatial	Meteorological + FWI	Excludes geographic coordinates (latitude, longitude) to disentangle spatial memorization from genuine meteorological predictive skill.

Table 4. Summarizes the definitions and justifications of the evaluation metrics used in this study.

Metric	Formula / Definition	Justification
Accuracy	$\frac{TP + TN}{(TP + TN + FP + FN)}$	Proportion of correctly classified observations among all predictions.
Precision	$\frac{\mathrm{TP}}{TP + FP}$	Proportion of predicted fire events that are true fires; reflects the false-alarm rate.
Recall	$\frac{\mathrm{TP}}{TP + FN}$	Proportion of actual fire events correctly detected; critical for minimizing missed fires.
F1-score	$\frac{2 . (\mathrm{Precision} \times \mathrm{Recall})}{\mathrm{Precision} + \mathrm{Recall}}$	Harmonic mean of Precision and Recall, providing a balanced performance measure.
ROC-AUC	Area under the ROC curve (TPR vs. FPR)	Measures the model's ability to discriminate between fire and non-fire cases across thresholds.
PR-AUC	Area under the Precision–Recall curve	Emphasizes performance on the positive class (fire events), particularly relevant for rare-event detection.

assessment of their individual and combined contributions to wildfire prediction performance.

Evaluation metrics

Six evaluation metrics are computed on the held-out test set for each of the four experimental configurations. Classification metrics (Accuracy, Precision, Recall, and F1-score) are obtained using a default probability threshold, while ranking metrics (ROC-AUC and PR-AUC) are computed from predicted probabilities (Table 4).

Post-hoc interpretability

To interpret the predictions of the best-performing model (XGBoost trained on the Full feature set), SHapley Additive exPlanations (SHAP) are computed using the shap.Explainer class with the predict_proba interface.

SHAP values quantify the marginal contribution of each feature to individual predictions relative to the expected model output, providing a locally accurate and globally consistent feature attribution framework.

The explainer is applied to a random subsample of 500 observations drawn from the test set using a fixed random seed. SHAP values were computed on this subsample (n = 500) to balance computational cost with representativeness.

The resulting SHAP values are visualized as a beeswarm plot conveying global feature importance and directional effects, a bar plot showing mean absolute SHAP values per feature, and a dependence plot for the top-ranked predictor. Group-level contributions are quantified by summing mean absolute SHAP values across the spatial, meteorological, and FWI feature groups.

RESULTS

Model performance and ablation study

Table 5 reports evaluation results for all configurations (two models × four feature sets) on the temporally held-out test set (2023–2024).

Table 5. Performance metrics for all model-feature configurations

Model	Feature Set	Accuracy	Precision	Recall	F1-score	ROC–AUC	ROC–AUC (CI)	PR–AUC	PR–AUC (CI)
XGBoost	Full	0.737	0.688	0.765	0.724	0.813	[0.800–0.824]	0.739	[0.717–0.761]
LightGBM	Full	0.735	0.683	0.770	0.724	0.810	[0.797–0.822]	0.736	[0.712–0.760]
XGBoost	No_FWI	0.731	0.684	0.754	0.717	0.804	[0.792–0.816]	0.728	[0.705–0.749]
LightGBM	No_FWI	0.726	0.677	0.750	0.712	0.803	[0.790–0.815]	0.724	[0.700–0.747]
XGBoost	No_meteo	0.730	0.673	0.783	0.724	0.805	[0.793–0.817]	0.726	[0.701–0.747]
LightGBM	No_meteo	0.724	0.668	0.773	0.717	0.804	[0.791–0.816]	0.722	[0.698–0.744]
XGBoost	No_spatial	0.666	0.622	0.661	0.641	0.733	[0.719–0.747]	0.657	[0.634–0.680]
LightGBM	No_spatial	0.662	0.613	0.683	0.646	0.734	[0.719–0.748]	0.652	[0.629–0.676]

Bootstrap confidence intervals for ROC-AUC and PR-AUC were estimated by resampling the test set 1,000 times with replacement; the 2.5th and 97.5th percentiles of the resulting distributions define the reported intervals. XGBoost Full achieves the highest ROC-AUC (0.813 [0.800–0.824]) and PR-AUC (0.739 [0.717–0.761]), while LightGBM Full attains marginally higher Recall (0.770). Both frameworks demonstrate near-equivalent performance (PR-AUC difference of 0.003), confirming that gradient boosting methods are equally capable with the complete 15-feature space. Among configurations including spatial features (Full, No_FWI, No_meteo), Precision ranges from 0.668 to 0.688, Recall from 0.750 to 0.783 and F1-scores from 0.712 to 0.724. The No_Spatial configuration yields substantially lower values (Precision: 0.613–0.622, Recall: 0.661–0.683), confirming the dominant contribution of geographic coordinates to predictive performance.

The ablation study demonstrates that FWI components significantly enhance predictive performance. Their removal (No_FWI) leads to a noticeable decline in both models (PR-AUC: –1.1 pp for XGBoost [0.739→0.728], –1.2 pp for LightGBM [0.736→0.724]; ROC-AUC: –0.9 pp and –0.7 pp), confirming that moisture codes (e.g., DC, DMC) capture critical multi-day fire risk dynamics beyond instantaneous weather conditions. Similarly, excluding meteorological variables (No_meteo) results in comparable performance degradation (PR-AUC: –1.3 pp for XGBoost, –1.4 pp for LightGBM), indicating that FWI and meteorological features provide complementary information. Notably, the No_Meteo configuration yields slightly higher Recall than the Full model (0.783 vs 0.770 for XGBoost), which reflects a precision-recall tradeoff: relying solely on FWI components produces a more permissive classifier that increases sensitivity at the cost of reduced precision (0.673 vs 0.688).

This behavior is consistent with the smoother, more integrated nature of FWI moisture codes, which tend to flag prolonged dry periods broadly, thereby capturing more fire events but also generating more false alarms.

The Full configuration consistently achieves the best results, highlighting the added value of combining both feature types. Finally, removing spatial features (No_spatial) causes the largest performance drop (≈ –8 pp in both ROC-AUC and PR-AUC), underscoring the strong influence of geographic context. Nevertheless, meteorological and FWI variables alone still maintain meaningful predictive power (ROC-AUC ≈ 0.733), confirming that the model captures genuine fire-weather relationships independent of location (Figure 3).

The confusion matrix for the XGBoost Full model (Figure 4) reveals 1.832 true negatives, 735 false positives, 498 false negatives, and 1.617 true positives. The LightGBM Full model yields comparable results with 1.810 true negatives, 757 false positives, 486 false negatives, and 1.629 true positives.

Feature importance and interpretability

Gain-based feature importance analysis (Figure 5) reveals that surface pressure (sp, 1983)

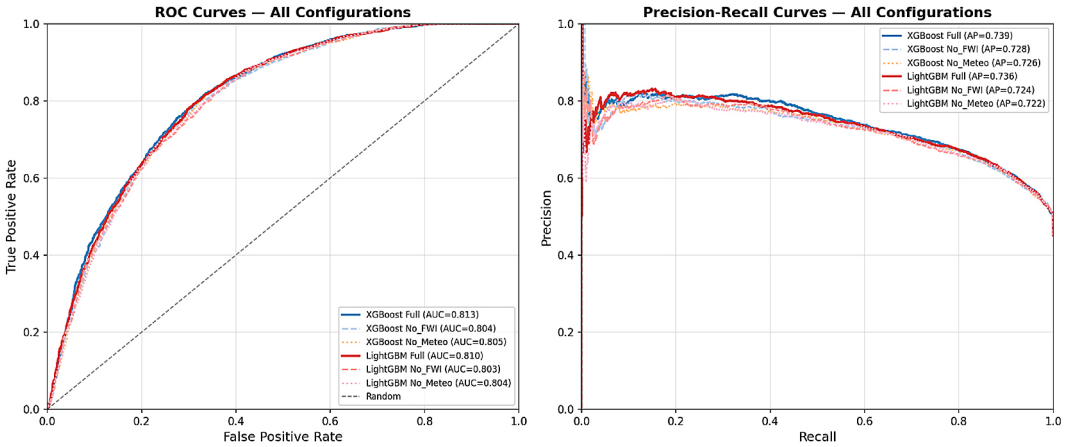

Figure 3. ROC curves (left) and precision-recall curves (right) for all model-feature configurations evaluated on the temporally held-out test set (2023–2024). The full configurations (solid lines) consistently outperform ablated variants across both metrics

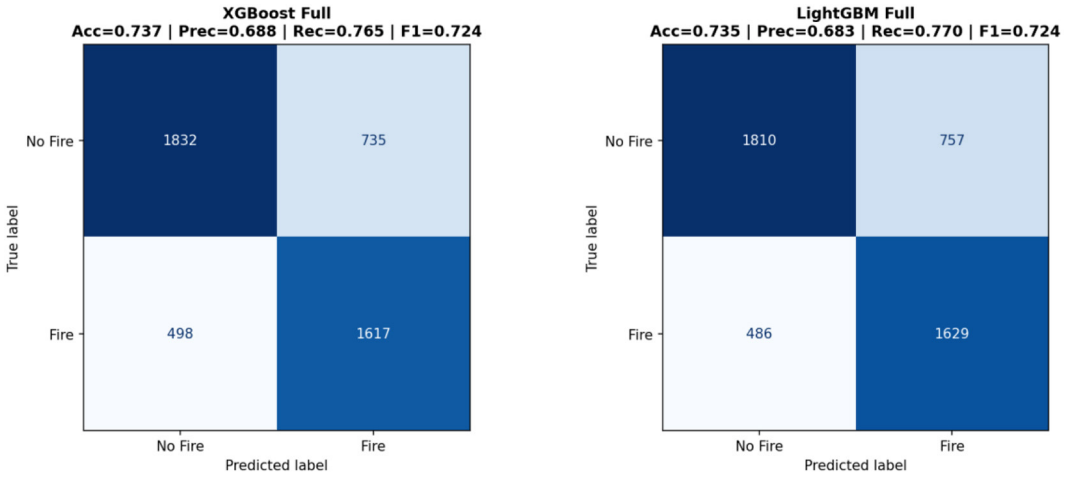

Figure 4. Confusion matrices for XGBoost full (left) and LightGBM full (right) on the held-out test set (2023–2024)

ranks first, followed by 2 m temperature (t2m, 1969) and surface solar radiation (ssr, 1823). High-pressure systems create extended dry periods conducive to fire, while temperature and radiation drive fuel moisture depletion. Relative humidity (RH, 1799) ranks fourth, followed by the drought code (DC, 1748), confirming the importance of both instantaneous atmospheric drying and long-term moisture deficit. Five of the top ten features are FWI components, with moisture codes (FFMC, DMC, DC) outranking the composite FWI index, suggesting that gradient boosting benefits from granular moisture state information rather than the aggregated danger rating. Spatial coordinates rank lower (longitude: 995, latitude: 912), consistent with the SHAP analysis which identifies longitude as the dominant predictor through a different attribution mechanism.

SHAP analysis

SHAP analysis on 500 test observations (Figures 6–8) confirms global feature importance and reveals consistent directional effects across predictors.

The SHAP bar plot (Figure 6) identifies longitude (mean |SHAP| = 0.18) and latitude (0.06) as dominant contributors, followed by key fire-weather variables including Drought Code (0.05), temperature (0.04), FFMC (0.04), and surface pressure (0.04). Group-level attribution shows

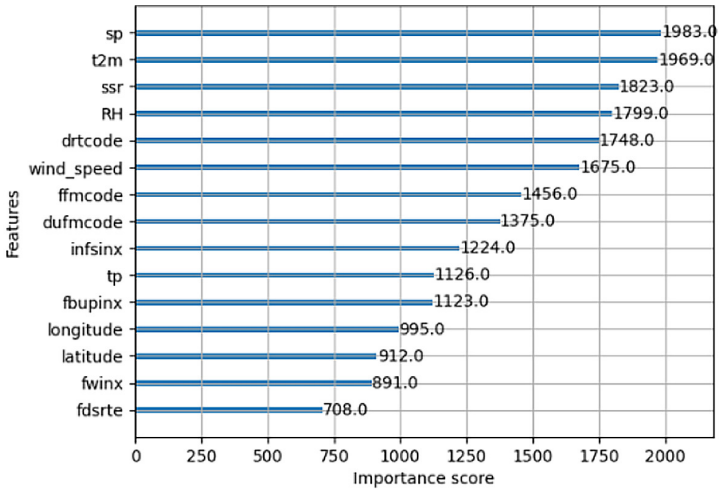

Figure 5. XGBoost feature importance (15 features). Scores represent average gain across 500 boosting rounds

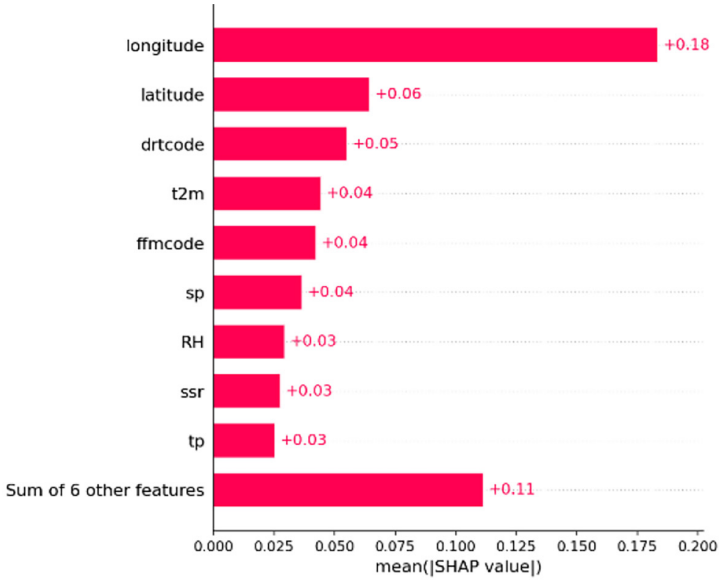

Figure 6. Mean absolute SHAP values per feature (XGBoost full)

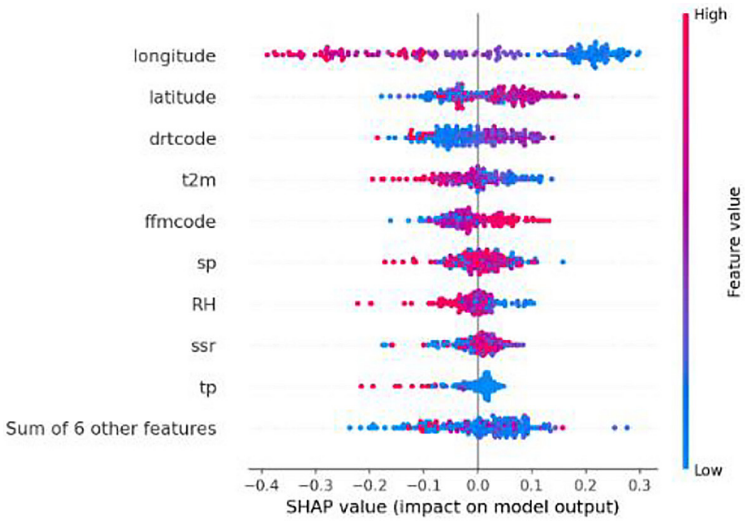

Figure 7. SHAP beeswarm plot showing directional feature effects (XGBoost full)

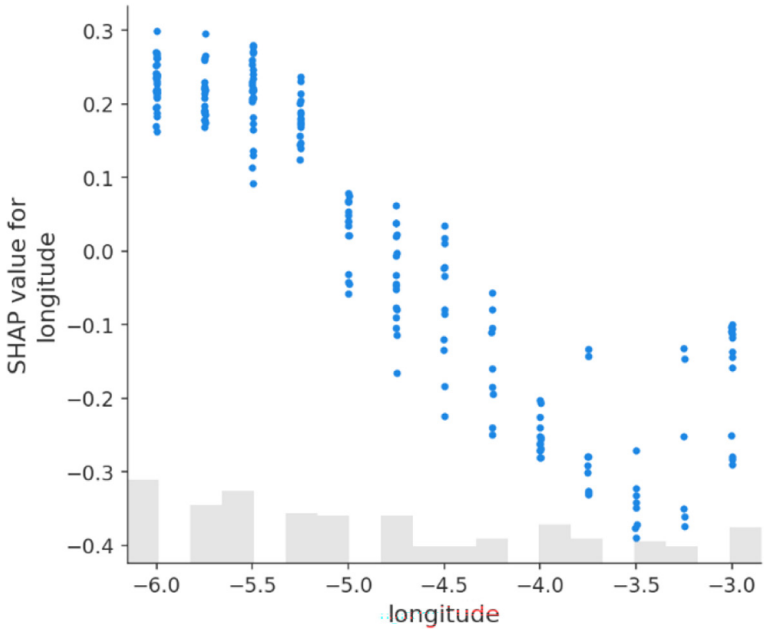

Figure 8. SHAP dependence plot for longitude (XGBoost full)

that spatial features (40.0%) slightly outweigh FWI components (31.5%) and meteorological variables (28.5%), highlighting the complementary role of all feature groups.

The beeswarm plot (Figure 7) illustrates clear physical relationships: high drought and temperature increase fire probability, while high precipitation reduces it. Surface pressure exhibits a mixed effect, reflecting more complex atmospheric influences.

The dependence analysis (Figure 8) reveals a strong west–east spatial gradient, with higher fire probability in western areas (Tétouan, Chefchaouen) and lower values in eastern regions (Al Hoceima). This pattern likely reflects underlying geographic and anthropogenic factors (e.g., vegetation, rainfall, human activity) captured indirectly through spatial coordinates, suggesting the need to include explicit explanatory variables in future work.

DISCUSSION

Performance and feature contributions

The best-performing model (XGBoost Full: PR-AUC = 0.739 [0.717–0.761], ROC-AUC = 0.813 [0.800–0.824]) demonstrates strong predictive capability for daily wildfire occurrence in northern Morocco. While higher accuracy has been reported in regional studies such as Seddouki et al. (2023), direct comparison remains limited due to differences in spatial scale, feature composition, and class balancing.

Moumane et al. (2025) applied XGBoost and LightGBM with SHAP to map wildfire susceptibility in the same TTA region, achieving accuracy of 0.920 using static vegetation and topographic variables (NDVI, LST, elevation). Mohajane et al. (2021) developed hybrid ML models for forest fire susceptibility in Northern Morocco, achieving AUC up to 0.989 using topographic and climatic variables. Unlike these susceptibility mapping approaches, which rely on static environmental variables, the present study focuses on dynamic daily meteorological predictability, making it more suitable for operational early warning systems. Regarding interpretability, Cilli et al. (2022) achieved an AUC of 0.813 using Random Forest with SHAP in a Mediterranean context, identifying FWI and NDVI as dominant predictors. Our study achieves comparable discriminative performance (ROC-AUC = 0.813) while excluding vegetation indices, confirming that meteorological and fire danger variables alone provide sufficient predictive skill for daily fire occurrence.

The ablation analysis confirms that FWI components and meteorological variables provide complementary information, capturing both multi-day fuel moisture dynamics and short-term atmospheric variability (No_FWI: –1.1 pp

PR-AUC; No_meteo: –1.3 pp PR-AUC). Spatial features contribute most strongly, with their removal causing the largest performance drop (–8.0 pp ROC-AUC; –8.2 pp PR-AUC), although fire-weather variables alone still retain meaningful predictive power (ROC-AUC = 0.733).

Feature importance analysis highlights surface pressure (importance = 1,983) as a key driver, reflecting the influence of persistent high-pressure systems on fire-conducive conditions. SHAP results further reveal a strong west–east spatial gradient (mean |SHAP| = 0.18), with higher fire probability in western provinces (Tétouan, Chefchaouen) compared to eastern areas (Al Hoceima). This pattern likely reflects underlying geographic and anthropogenic factors captured indirectly through spatial coordinates, rather than a direct causal effect, emphasizing the need for explicit human-related variables in future models.

Operational implications

The model achieves a Recall of approximately 0.765–0.770, enabling effective detection of most fire events and supporting early intervention strategies. Precision remains moderate (0.683–0.688), reflecting the inherent trade-off between sensitivity and specificity under the balanced evaluation framework. The use of probabilistic outputs allows flexible threshold adjustment based on operational needs. Integration into systems such as Copernicus EFFIS could enhance decision-making by combining predictive modeling with real-time monitoring.

To further assess the robustness of the proposed framework under real-world conditions, an additional evaluation was conducted using a test set with the natural class distribution (~2.78% fire occurrence). Under this setting, precision decreases substantially (≈0.07), while recall remains high (≈0.75), which is expected in rare-event detection problems where models prioritize sensitivity over specificity. The achieved PR-AUC (~0.10) remains significantly higher than the baseline defined by fire prevalence (~0.03), indicating that the model retains meaningful predictive skill under operational conditions. These findings highlight the intrinsic difficulty of wildfire prediction under severe class imbalance and emphasize the importance of threshold calibration when deploying such models in real-world early warning systems.

Limitations

Several limitations constrain generalization. First, the absence of anthropogenic variables is critical, as most Mediterranean fires are human-induced (Vilar et al., 2016; Balch et al., 2017); spatial coordinates act only as indirect proxies. Second, the coarse spatial resolution (0.25°, ~784 km² per cell) and satellite detection limits (MODIS ~50 MW threshold) bias the dataset toward larger fires. Third, spatial autocorrelation between neighboring grid cells may inflate performance estimates, as temporal validation does not ensure spatial independence; future work should implement spatial blocking by subregion. Fourth, the fire weather index was originally developed for Canadian boreal ecosystems (Van Wagner, 1987), introducing uncertainty in Mediterranean contexts dominated by cork oak, matorral, and Moroccan fir.

Specifically, the moisture codes (FFMC, DMC, DC) were calibrated for boreal fuel types and may not accurately represent the drying dynamics of Mediterranean vegetation such as cork oak, Thuya and maquis shrublands, which exhibit different fuel moisture retention characteristics and ignition thresholds.

As a result, the FFMC and DMC thresholds associated with high ignition probability in Canadian boreal forests may not translate directly to the TTA region, potentially introducing systematic bias in danger rating estimation particularly during transitional seasons (spring and autumn) when Mediterranean vegetation undergoes rapid moisture fluctuations distinct from boreal drying patterns.

Finally, key environmental factors such as vegetation dynamics (NDVI), fuel load, and topography are not included.

CONCLUSIONS

This study addressed a critical gap in wildfire prediction research for Mediterranean North Africa by developing and rigorously evaluating a machine learning framework for the Tangier–Tétouan–Al Hoceima region. By adopting a temporal hold-out evaluation strategy and a structured ablation design, this work provides the first benchmark performance metrics for daily fire occurrence prediction in the TTA region

under temporally consistent validation conditions (ROC-AUC = 0.813, PR-AUC = 0.739).

Three principal findings emerge from this work. First, the joint integration of ERA5 meteorological variables with FWI components consistently outperforms either group alone, confirming that instantaneous atmospheric conditions and multi-day fuel moisture dynamics capture complementary and non-redundant aspects of fire risk. This finding supports the continued operational use of ERA5-derived FWI in Mediterranean fire danger forecasting systems such as EFFIS, despite the known calibration mismatch between Canadian boreal fuel assumptions and Mediterranean vegetation types.

Second, the spatial sensitivity analysis reveals that geographic coordinates contribute substantially to predictive performance (~8 pp ROC-AUC), yet meteorological and FWI variables alone retain meaningful discriminative ability (ROC-AUC = 0.733). This distinction confirms that the model captures genuine fire-weather relationships independent of geographic memorization, while the strong spatial contribution reflects the complex interplay of vegetation distribution, orographic gradients, and anthropogenic pressure that characterize the TTA region.

Third, SHAP analysis identifies a dominant west-to-east spatial gradient that the model captures through longitude as a composite proxy for multiple co-varying geographic factors population density, road networks, vegetation distribution, and orographic rainfall gradients. This finding underscores a fundamental limitation shared by purely meteorological fire prediction frameworks: the absence of explicit anthropogenic ignition drivers, which account for approximately 95% of Mediterranean fire starts.

The reproducible and interpretable framework developed here demonstrates that ERA5 meteorological variables and FWI components provide complementary and non-redundant predictive information for daily wildfire occurrence in the TTA region. The retained predictive skill of purely meteorological configurations (ROC-AUC = 0.733) confirms that fire-weather relationships captured by the model extend beyond spatial memorization, while the identified west-to-east spatial gradient highlights the role of geographic and anthropogenic factors in structuring regional fire risk.

REFERENCES

1. Abatzoglou J.T., Kolden C.A., Cullen A.C., Sadegh M., Williams E.L., Turco M., Jones M.W. (2025). Climate change has increased the odds of extreme regional forest fire years globally. *Nature Communications*, *16*, 6390. https://doi.org/10.1038/s41467-025-61608-1
2. Balch J.K., Bradley B.A., Abatzoglou J.T., Nagy R.C., Fusco E.J., Mahood A.L. (2017). Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences*, *114*(11), 2946–2951. https://doi.org/10.1073/pnas.1617394114
3. Boubekraoui H., Maouni Y., Ghallab A., Draoui M., Maouni A. (2023). Wildfires risk assessment using hotspot analysis and results application to wildfires strategic response in the region of Tangier-Tetouan-Al Hoceima, Morocco. *Fire, 6*(8), 314. https://doi.org/10.3390/fire6080314
4. Castro I., Stan A.B., Taiqui L., Schiefer E., Ghallab A., Derak M., Fulé P.Z. (2022). Detecting fire-caused forest loss in a Moroccan protected area. *Fire*, *5*(2), 51. https://doi.org/10.3390/fire5020051
5. Cilli R., Elia M., D'Este M., Giannico V., Amoroso N., Lombardi A., Pantaleo U., Monaco A., Tangaro S., Bellotti R., Lafortezza R. (2022). Explainable artificial intelligence (XAI) detects wildfire occurrence in the Mediterranean countries of Southern Europe. *Scientific Reports*, *12*, 16349. https://doi.org/10.1038/s41598-022-20347-9
6. Di Giuseppe F., Vitolo C., Krzeminski B., Barnard C., Maciel P., San-Miguel J. (2020). ERA5-based global meteorological wildfire danger maps. *Scientific Data*, *7*, 216. https://doi.org/10.1038/s41597-020-0554-z
7. Driouech, F., et al. (2021). Recent observed country-wide climate trends in Morocco. *International Journal of Climatology.* https://doi.org/10.1002/joc.6734
8. Hersbach H., Bell B., Berrisford P., Hirahara S., Horányi A., Muñoz-Sabater J., Thépaut J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803
9. Iban M.C., Sekertekin A. (2022). Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces, Turkey. *Ecological Informatics*, *69*, 101647. https://doi.org/10.1016/j.ecoinf.2022.101647
10. Jain P., Coogan S.C.P., Subramanian S.G., Crowley M., Taylor S., Flannigan M.D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, *28*(4), 478–505. https://doi.org/10.1139/er-2020-0019
11. Jones M.W., Abatzoglou J.T., Veraverbeke S., Andela N., Lasslop G., Forkel M., Smith A.J.P., Burton C., Betts R.A., van der Werf G.R., Sitch S.,

Canadell J.G., Santín C., Kolden C., Doerr S.H., Le Quéré C. (2022). Global and regional trends and drivers of fire under climate change. *Reviews of Geophysics*, *60*(3), e2020RG000726. https://doi.org/10.1029/2020RG000726

12. Lundberg S.M., Lee S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.
13. Maranón T., Ajbilou R., Ojeda F., Arroyo J. (2004). Ecological and biogeographical analyses of Mediterranean forests of northern Morocco. *Acta Oecologica*, *25*(1–2), 1–9. https://doi.org/10.1016/j.actao.2005.08.006
14. Mohajane M., Costache R., Firoozeh Karimi, Quoc Bao Pham, Ali Essahlaoui, Hoang Nguyen, Giovanni Laneve, Fatiha Oudija, Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area, *Ecological Indicators*, https://doi.org/10.1016/j.ecolind.2021.107869
15. Moumane A., Al Karkouri A., Elmotawakkil A., Alkhuraiji W.S., Rebouh N.Y. and Youssef Y.M. (2025) Advancing wildfire susceptibility mapping through machine learning and SHapley Additive exPlanations-integrated geospatial analysis in Northern Morocco's Mediterranean region. *Front. For. Glob. Change 8*, 1705341. https://doi: 10.3389/ffgc.2025.1705341
16. Moghim, S., Mehrabi, M. (2024). Wildfire assessment using machine learning algorithms in different regions. *Fire Ecology*, *20*, 117. https://doi.org/10.1186/s42408-024-00335-2
17. Salhi, A., et al. (2019). Rainfall distribution and trends of the daily precipitation concentration index in northern Morocco. *SN Applied Sciences, 1*, 277. https://doi.org/10.1007/s42452-019-0290-1
18. Seddouki M., Benayad M., Aamir Z., Tahiri M., Maanan M., Rhinane H. (2023). Using machine learning coupled with remote sensing for forest fire susceptibility mapping. Case study Tetouan province, Northern Morocco. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*(4/W6), 333–342. https://doi.org/10.5194/isprs-archives-XLVIII-4-W6-2022-333-2023
19. Sjöström J., Vermina Plathner F., Granström A. (2025). 70 years of observational weather data show increasing fire danger for boreal Europe and reveal bias of ERA5 reanalysed data. *Scientific Reports*, *15*, 20111. https://doi.org/10.1038/s41598-025-04200-3
20. Van Wagner C.E. (1987). *Development and structure of the Canadian Forest Fire Weather Index System* (Forestry Technical Report 35). Canadian Forestry Service, Ottawa.
21. Vilar L., Woolford D.G., Martell D.L., Martín M.P. (2016). A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. *International Journal of Wildland Fire*, *19*(3), 325–337. https://doi.org/10.1071/WF09030
22. Xu Z., Li J., Cheng S., Rui X., Zhao Y., He H., Guan H., Sharma A., Erxleben M., Chang R., Xu L. (2025). Deep learning for wildfire risk prediction: integrating remote sensing and environmental data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *227*, 632–677. https://doi.org/10.1016/j.isprsjprs.2025.06.002
23. Zaidi A. (2023). Predicting wildfires in Algerian forests using machine learning models. *Heliyon*, *9*(7), e18064. https://doi.org/10.1016/j.heliyon.2023.e18064