Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning Prediction of Early-Season Wildfire Risk in Ghana's Guinea Savannah Using Multi-Source Data and CMIP6 Climate Scenarios

Wallelgn melese tema

hanielcj1234@gmail.com

pan African university for water and energy science https://orcid.org/0009-0006-2127-7609

Research Article

Keywords: wildfire risk, machine learning, Ghana, Guinea Savannah, CMIP6, feature importance, operational prevalence, fire weather, climate change

Posted Date: July 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10192384/v1

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

Additional Declarations: The authors declare no competing interests.

Machine Learning Prediction of Early-Season Wildfire Risk in Ghana's Guinea Savannah Using Multi-Source Data and CMIP6 Climate Scenarios

Author: Wallelgn Tema

Institute: Pan African University Institute of Water and Energy Sciences (PAUWES), University of Tlemcen

Country and city: Algeria, Tlemcen

Contact author: hanielcj1234@gmail.com

Abstract

Wildfires in Ghana's Guinea Savannah zone affect large areas each dry season, yet Ghana still lacks a locally calibrated early-season fire danger tool for the northern savannah. This study develops and evaluates a machine learning fire-risk prototype for November-January using a 1 km gridded dataset for 2001-2024. The dataset combines MODIS burned area, NASA FIRMS active fire detections, ERA5-Land, CHIRPS, FLDAS soil moisture, MODIS vegetation products, SRTM topography, and human-access variables. Twenty-one base predictor variables were expanded into 123 model features using monthly lags, rolling statistics, anomalies, drought indices, and seasonal indicators. Five algorithms were compared: Random Forest, XGBoost, LightGBM, CatBoost, and a multilayer perceptron. On the stratified test sample, the tree-based models achieved high discrimination, with AUC values of 0.960-0.962. Because the model-development sample was deliberately balanced relative to the full cell-month population, threshold-dependent metrics such as precision, F1, MCC, and Brier score are treated as sampling-dependent and are interpreted with prevalence-adjusted estimates. The full-feature LightGBM model exceeded the monthly aggregated Canadian Fire Weather Index by an absolute AUC difference of 0.28, although this comparison reflects both algorithm choice and the richer feature set rather than machine learning alone. Feature attribution did not support a single dominant soil-moisture control. Instead, temperature range, population density, surface pressure, precipitation memory, vegetation state, and soil-moisture memory all contributed to predicted fire risk. Surface pressure and other static or near-static variables are interpreted cautiously because they may encode spatial location and elevation. Bias-corrected CMIP6 experiments suggest higher late-century fire risk under SSP2-4.5 and SSP5-8.5, but these projections are presented as illustrative because tree ensembles cannot extrapolate beyond the training envelope and human drivers were held constant. The revised index should therefore be viewed as a research prototype that can support future operational testing, not as a fully validated warning system.

Keywords: wildfire risk; machine learning; Ghana; Guinea Savannah; CMIP6; feature importance; operational prevalence; fire weather; climate change

1 Introduction

1.1 Background and Motivation

Wildfires are a recurring phenomenon in West African savannah ecosystems, with significant ecological and socioeconomic impacts that extend far beyond the immediate fire-affected areas [1], [2]. Every year, about 2 to 3 million hectares of land in Ghana burn, mostly in the Guinea Savannah zone during the dry season (November to March) [3]. These fires affect agricultural productivity, biodiversity conservation, carbon stocks, and rural livelihoods, creating cascading effects on food security, ecosystem services, and regional climate patterns [4].

The Guinea Savannah ecological zone in northern Ghana is characterized by fire-adapted vegetation and a long history of anthropogenic burning for land clearing, hunting, and pasture management [5]. This region represents a complex socio-ecological system where fire regimes are shaped by the interplay of climatic conditions, vegetation dynamics, and human decision-making [6]. Understanding and predicting fire occurrence in such systems requires integrating diverse data sources and capturing non-linear interactions across multiple scales.

The November-January period is important for fire management because it captures the start and peak of the dry-season fire cycle in this dataset. In the fire record analysed here, November, December, and January account for more than 87% of detections. Earlier studies often describe early dry-season burns as less intense and more controllable than late-season fires, but this manuscript does not include fire-intensity observations. For that reason, the term early-season in this study refers to the target prediction window rather than to a demonstrated reduction in intensity.

1.2 Current State of Fire Danger Indices

Fire danger indices are important tools for predicting and controlling the risk of wildfires. The Canadian Fire Weather Index (FWI) System [8] is the most widely used fire danger rating system globally, integrating daily weather observations of temperature, relative humidity, wind speed, and precipitation to produce numerical ratings of fire danger. FWI has been applied in various tropical and subtropical regions, including parts of Africa [9].

However, FWI was developed for Canadian boreal forests and may not capture the specific dynamics of grass-dominated fuel systems, human ignition patterns, and seasonal moisture variations characteristic of West African savannahs [9]. Specific limitations include the following:

1. Fuel characteristics mismatch: FWI assume continuous forest fuel beds, while savannas have patchy grass-dominated fuels that respond differently to moisture changes [6].

2. Sparse meteorological coverage: Ghana has limited meteorological station density, making spatially explicit FWI calculations challenging [10].

3. Ignition source omission: FWI does not account for anthropogenic ignitions, which

account for over 90% of fires in West Africa [6].

4. Live fuel moisture neglect: FWI ignores live fuel moisture dynamics, which are critical in savannah ecosystems where grasses cure seasonally [11].

5. Soil moisture absence: FWI does not incorporate soil moisture, which is an important memory variable for fuel drying and vegetation curing in the models tested here.

Recent work in African savannahs shows that standard weather-only indices can miss important local controls on fire occurrence. This supports the development of regionally calibrated models that include climate, fuel, moisture, and human-access variables.

1.3 Machine Learning Approaches for Fire Prediction

Machine learning approaches offer promising alternatives by integrating diverse data sources and capturing complex nonlinear relationships [13]. These methods can automatically learn the hierarchical interactions between climate, vegetation, topography, and human factors that determine fire occurrences without requiring a priori assumptions about the functional form of these relationships.

Random Forest [14] and XGBoost [15] have been successfully applied for fire prediction globally, demonstrating superior performance compared to traditional indices. [13] used machine learning with CMIP6 constraints to project elevated global socioeconomic risks from wildfire, finding that western and central Africa face among the highest future risk increases. This points out the importance of region-specific tools that can support adaptation planning. In Ghana, foundational work by [3] and [16] has established baseline fire dynamics using Google Earth Engine. [16] analysed environmental and climatic changes contributing to wildfire spread, identifying key drivers including precipitation, evapotranspiration, temperature, humidity, and soil moisture. [3] documented fire trends in the Guinea Savannah and forest-savannah mosaic zones from 2001 to 2021. Most recently, [3] applied Random Forest with CMIP6 to project future fire spread under climate change.

2. Materials and Methods

2.1 Study Area

The study area encompasses the Guinea Savannah ecological zone of northern Ghana, extending from approximately 7.5°N to 11.5°N latitude and -3.5°E to 1.5°E longitude (Figure 1). This zone covers approximately 147,900 km² across the Upper East, Upper West, Northern, and Northeastern regions Table 1 Study area characteristics of the Guinea Savannah zone, Ghana.. The region was selected based on its high fire activity, with 2-3 million hectares burning annually, and its ecological and socioeconomic importance to Ghana **Figure** 1 Study **Area map**shows the study area map and characteristics of the study Area

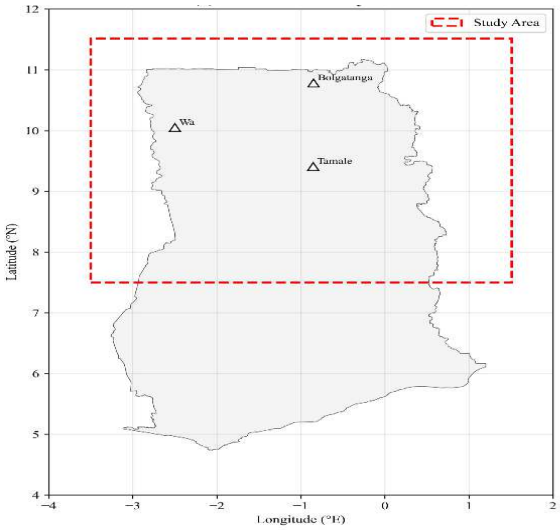


Figure 1 Study Area map

Table 1 Study area characteristics of the Guinea Savannah zone, Ghana.

Parameter	**Value**
Total Grid Points	200,000
Points ≥7.5°N	200,000
Latitude Range	7.51° – 11.50° N
Longitude Range	-3.49° – 1.50° E
Total Area	147,900 km²

The climate is tropical wet-dry (Aw in the Köppen classification) with a single rainy season from May to October and a prolonged dry season from November to April. Mean annual rainfall ranges from 800 to 1200 mm, with over 90% falling between May and October [10]. Mean annual temperatures range from 26-30°C, with maximum temperatures reaching 35-40°C in March-April before the rains.The vegetation is fire-adapted Guinea savannah woodland, characterised by a continuous grass layer dominated by perennial grasses such as *Andropogon gayanus* and *Hyparrhenia* species and scattered fire-tolerant trees, including shea (*Vitellaria paradoxa*), dawadawa (*Parkia biglobosa*), and *Anogeissus leiocarpa* [17]. Tree cover ranges from 5 to 30%, decreasing with increasing fire frequency. Fire regimes in the region are characterised by annual fires during the dry season, with two distinct periods: early dry season (November-January), when fires are typically less intense and patchier, and late dry season (February-March), when fires are more severe and extensive [16]. Approximately 70% of the zone burns annually, with significant interannual variability driven by rainfall patterns. Over 90% of fires are anthropogenic in origin, set for land clearing, hunting, pasture management, and pest control [6].

2.2 Data Acquisition and Preprocessing

A multi-source dataset for January 2001–December 2024 (288 months) was harmonized to a 1 km grid. Fire occurrence combined MODIS MCD64A1 burnt area and NASA FIRMS active fire data [18,13]. Climate came from CHIRPS precipitation [19,20] and ERA5-Land temperature, dewpoint, pressure, and wind [21], used to compute monthly temperature, humidity, wind, and temperature range. Soil moisture from FLDAS for four layers (0–10 cm, 10–40 cm, etc.) was interpolated to 1 km [13,22]. Vegetation indices (NDVI, EVI, LAI, FPAR) were derived from MODIS [23,24], topography from SRTM [25], and human factors from World Pop, GRIP, GHSL, and nighttime lights [26,27]. Drought indices SPI and SPEI were calculated from Terra Climate precipitation and PET at 3-, 6-, and 12-month scales [28,29]..

2.3 Feature Engineering

To capture temporal dynamics crucial for early-season prediction, we engineered additional features from the base monthly data. This approach follows best practices from [13] and [30].

2.3.1 Lag Variables

For all climate, soil moisture, and vegetation variables, we created lagged versions at 1, 2, 3, 4, 8, and 12 months to capture antecedent conditions. These lags were selected based on:

1-4 months: Recent weather effects on fuel moisture and curing

8 months: Medium-term drought accumulation and vegetation response

12 months: Seasonal memory effects and carry-over from previous fire seasons

2.3.2 Rolling Statistics

For precipitation and temperature variables, we calculated rolling statistics over windows of 3 months, 6 months, and 12 months:

Rolling mean: Captures average conditions over the window

Rolling maximum: Captures extreme events that may influence fuel production

Rolling minimum: Captures critical low points for fuel moisture

Rolling standard deviation: Captures variability as an indicator of instability

2.3.3 Climate Anomalies

For temperature, precipitation, and soil moisture, we calculated anomalies relative to the 2001-2020 monthly climatology:

Absolute anomaly: $X_{anom} = X_{obs} - \bar{X}_{clim}$

Standardized anomaly: $X_{std} = (X_{obs} - \bar{X}_{clim})/\sigma_{clim}$

These anomalies capture deviations from normal conditions, which are often more predictive than absolute values for extreme events.

2.3.4 Seasonal Indicators

Binary categorical variables were added for:

Month (1-12)

Season: early dry (November-January), late dry (February-March), wet (April-October)

Week of year (1-52)

The feature engineering process resulted in 123 model features from 21 base predictor variables for each monthly grid-cell observation. Table 2 summarises the feature groups used in the model-development dataset.

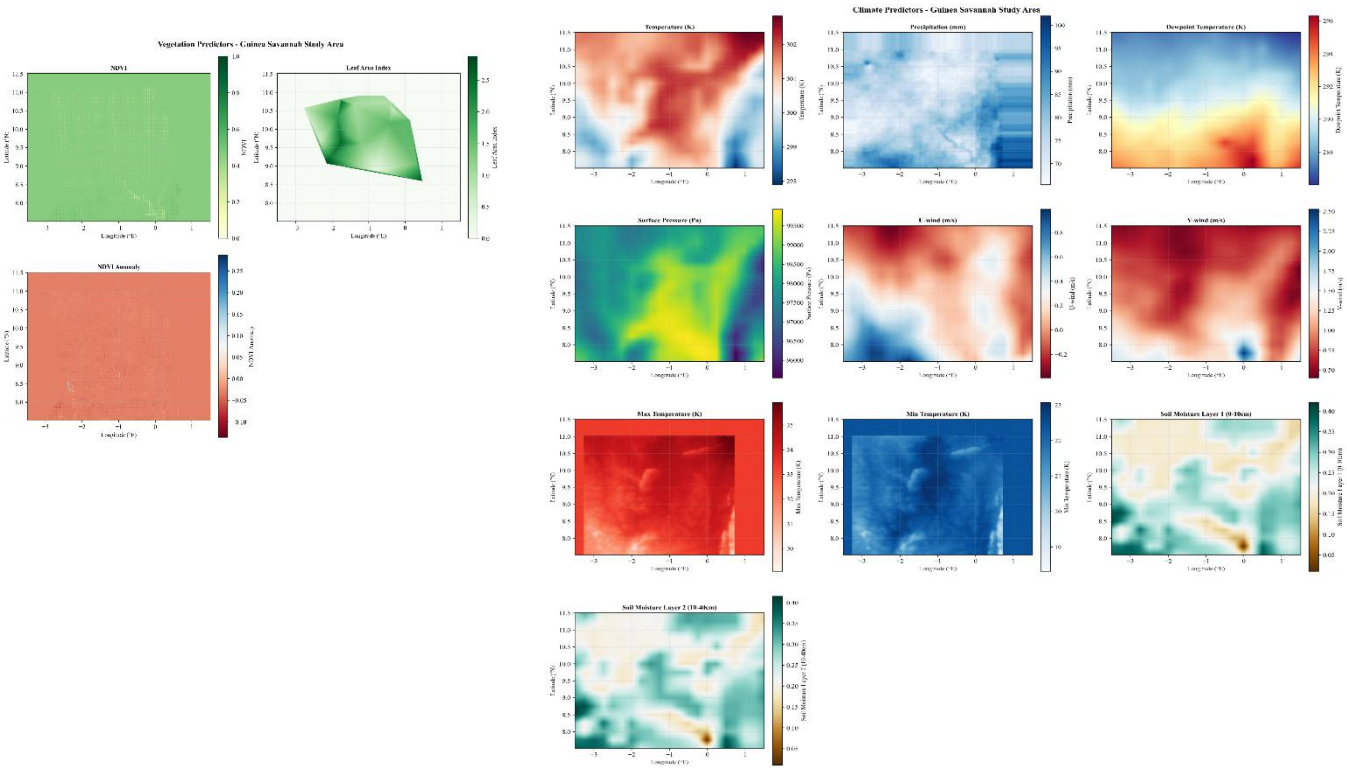
Table 2 Summary of feature categories and counts in the final dataset.

Feature Category	Count
Base environmental variables	16
Derived variables (wind speed, VPD, temperature range)	5
FWI components	6
Engineered lag, rolling, anomaly, and seasonal features	96
Total model features	**123**

Figure 2, Figure 3, and Figure 4 illustrate the climate and vegetation predictors, human and static predictors, and CMIP6 climate projection inputs used in this research.

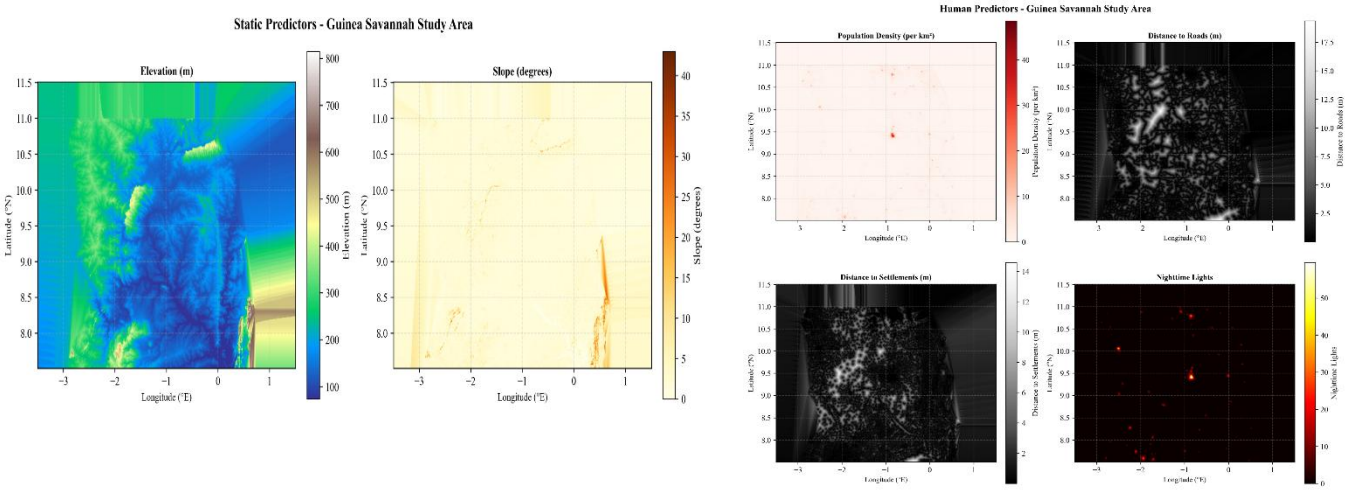
Climate And Vegetation Predictor Dataset

Figure 2 Climate and Vegetation Predictor Dataset



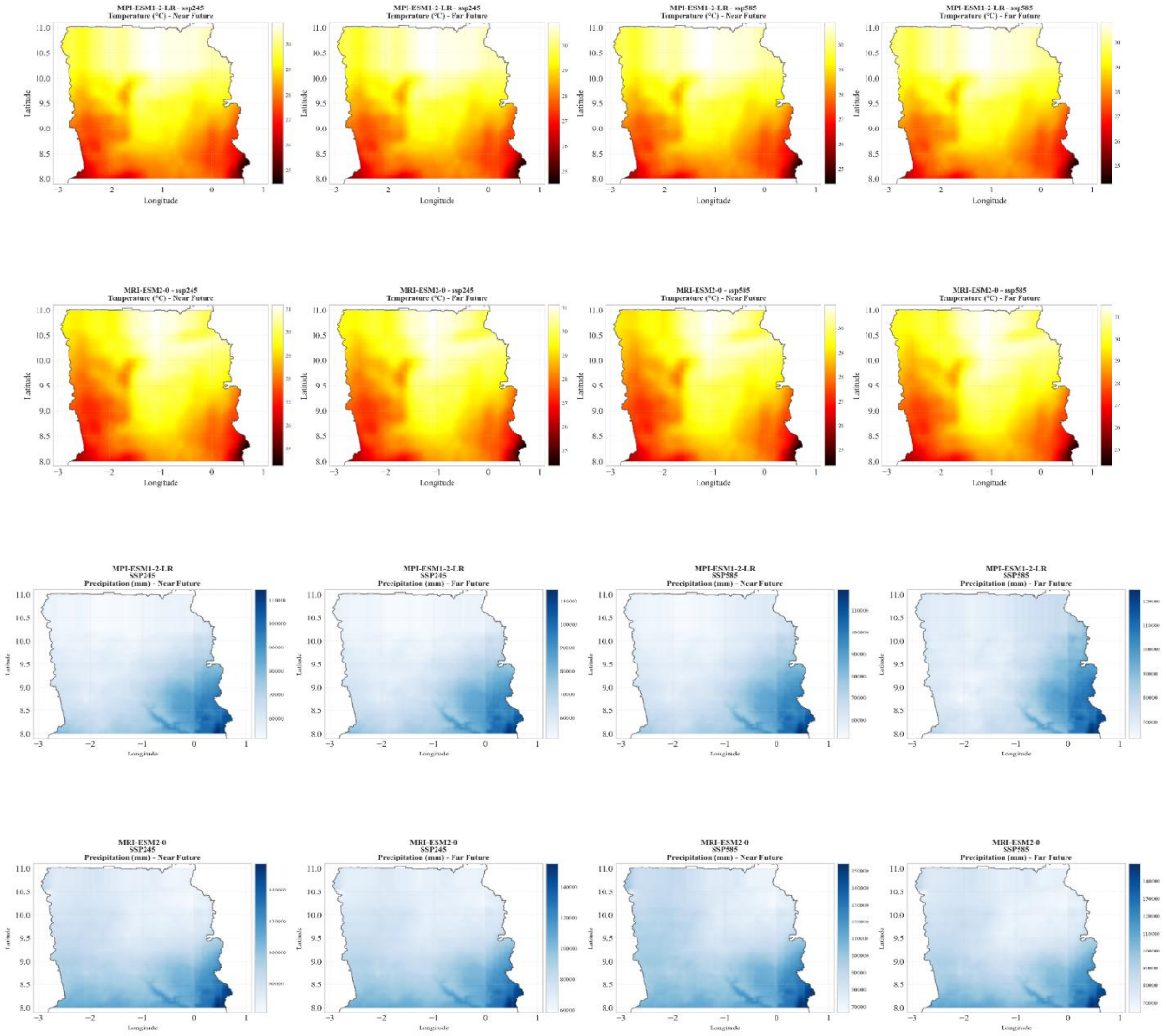
Human And Static Predictor Dataset

Figure 3 Human and Static Predictor Dataset



CMIP6 Climate Projection Dataset

Figure 4 CMIP6 Future predication



2.4 Data Sampling and Temporal Split

Table 3 reports the stratified model-development sample and should not be interpreted as the full operational fire prevalence of the 1 km monthly grid.

Dataset	Samples	Fire Rate (%)
Training	336,674	33.94
Validation	62,393	33.65
Test	100,931	31.10
Total	**499,998**	**33.33**
Full gridded cube (before sampling)	≈57.6 million	≈1.7*

Table 3 Dataset characteristics and temporal split for the stratified model-development sample. The full gridded cube contains approximately 57.6 million cell-month observations (about 200,000 grid cells over 288 months). The 499,998 observations used for model development represent a stratified sample with fire prevalence close to one positive observation for every two negative observations. This design makes model training tractable and reduces class imbalance, but it changes the base rate. AUC is therefore the main prevalence-independent discrimination metric, while precision, F1, MCC, and Brier score are reported as sample-conditioned metrics unless they are explicitly reweighted to the operational prevalence.

2.5 Machine Learning Models

Five machine learning algorithms were implemented and compared, including Random Forest [14], XGBoost [32], LightGBM [33], CatBoost [34], and a Multi-Layer Perceptron (MLP) neural network. All models were trained on the 2001–2016 set, tuned on 2017–2019, and evaluated on 2020–2024. Random Forest used 300 trees, max depth 20, min samples split 10, min leaf 5, and balanced class weights. XGBoost used 500 estimators, depth 8, learning rate 0.05, subsample/colsample 0.8, and scale positive weight equal to the negative-to-positive ratio. LightGBM employed 500 estimators, depth 10, 31 leaves, learning rate 0.05, and scale positive weight adjusted for class imbalance. CatBoost used 500 iterations, depth 8, learning rate 0.05, L2 leaf 3, and border count 128. The MLP had four hidden layers (256, 128, 64, 32) with ReLU, input 123 features, output 2 neurons with softmax, trained with Adam (lr 0.001), batch size 256, up to 200 epochs with early stopping, and class weights, using a 50,000-sample subset for computational efficiency.

2.6 Hyperparameter Tuning and Validation

Hyperparameter tuning was performed using Optuna [31] with 100 trials per model and a tree-structured Parzen estimator sampler. To reduce, but not eliminate, spatial leakage during tuning, 5-fold spatial cross-validation was implemented by clustering grid-cell coordinates into spatial groups. Models were trained on four groups and evaluated on the held-out group, and the configuration with the highest mean validation AUC was selected. The headline test scores reported below come from the temporal holdout period, not from the spatial cross-validation. Because k-means spatial grouping can still permit autocorrelation leakage along cluster boundaries, a future version should use buffered or block cross-validation as a stricter spatial test.

2.7 Model Interpretability with SHAP

SHapley Additive explanations (SHAP) [36] were used to interpret model predictions, quantifying each feature's contribution to individual predictions (positive values increase, negative decrease fire probability) based on cooperative game theory.

For each prediction, the SHAP decomposition is written as:

$$f(x) = \varphi 0 + \Sigma j=1M\ \varphi j$$

where $\varphi 0$ is the baseline model output and φj is the contribution of feature j. SHAP was used to estimate global feature importance from mean absolute SHAP values, generate summary and dependence plots, identify interactions, and explain individual predictions. TreeExplainer was used for the tree-based models. For the MLP, first-layer weights were used only as an exploratory proxy for feature importance and are not directly comparable with SHAP values.

2.8 Future Climate Projections

Future climate projections were derived from two CMIP6 Earth System Models, MPI-ESM1-2-LR [38] and MRI-ESM2-0 [39], under SSP2-4.5 and SSP5-8.5 [37,40] for the near future (2040-2059) and far future (2080-2099). Raw outputs were bias-corrected using quantile mapping [41,42] with ERA5-Land and CHIRPS for 2001-2020 as observational references. Corrected daily data were aggregated to monthly predictors and processed with the same historical workflow. Static predictors, including topography, population, roads, and settlements, were held constant. This assumption is a first-order limitation because human-access variables are important predictors in the historical model. In addition, tree ensembles cannot extrapolate smoothly beyond the range of their training data. Projected changes are therefore reported as approximate ranges and should be read as scenario-based sensitivity experiments rather than precise forecasts.

2.9 Performance Evaluation Metrics

Model performance was evaluated using multiple metrics to capture different aspects of predictive skill:

- Area Under the ROC Curve (AUC-ROC): discrimination between fire and non-fire observations across thresholds. AUC is independent of class prevalence, making it the most robust metric for the stratified sample used here.

- F1 score: harmonic mean of precision and recall. Because it depends on the number of predicted positives and observed positives, it changes when the evaluation prevalence changes.

$$F1 = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$$

Precision: proportion of predicted fire observations that were observed as fire.

$$Precision = \frac{TP}{TP+FP}$$

Recall: proportion of observed fire observations that were correctly predicted as fire.

$$Recall = \frac{TP}{TP + FN}$$

Matthews Correlation Coefficient (MCC): balanced summary of the full confusion matrix.

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{((TP + FP)(TP + FN)(TN + FP)(TN + FN))}}$$

Brier score: mean squared error of predicted probabilities. Operational calibration must be evaluated at the operational base rate, not only on the stratified sample.

$$BS = (1/N)\ \Sigma i=1N\ (pi - oi)^2$$

The Canadian Fire Weather Index (FWI) was calculated following the original Canadian FWI system [8] using ERA5-Land weather inputs. In this study, FWI is used as a weather-only baseline. The comparison with full-feature machine learning models should therefore be interpreted as a comparison between a weather-only index and models that also include soil moisture, vegetation, topography, and human-access information. A stricter future benchmark should compare native daily FWI with a weather-only machine learning model at the same temporal resolution.

3. Results

3.1 Fire Occurrence Patterns and Seasonality

Analysis of the 288-month fire record (2001-2024) revealed strong seasonality in fire occurrence across the Guinea Savannah zone, consistent with the region's distinct wet and dry seasons. A total of 971,529 fire events were recorded during the study period, with a clear peak during the dry season months **(Table 4,).**

Table 4 Monthly fire occurrence distribution (2001-2024).

Month	Fire Count	Percentage
January	313,162	32.2%
February	87,114	9.0%
March	21,572	2.2%
April	6,906	0.7%
May	991	0.1%
June	107	0.01%
July	253	0.03%
August	727	0.07%
September	213	0.02%
October	2,419	0.2%
November	119,934	12.3%
December	418,131	43.0%
Total	**971,529**	**100%**

December recorded the highest number of fires (418,131), accounting for 43% of all fire events, followed by January (313,162, 32.2%) and November (119,934, 12.3%). Together, these three months account for over 87% of all fire activity. The concentration of fires in the early- to mid-dry season reflects the combination of sufficient grass fuel accumulation from the preceding wet season and the onset of dry conditions that make this fuel flammable.

February and March, representing the late dry season, show declining fire activity (9.0% and 2.2%, respectively), likely due to fuel depletion rather than lack of ignitions. By late dry season, much of the available grass fuel has already been consumed, limiting further fire spread even though conditions remain favorable.

The wet season months (June-September) show minimal fire activity, with fewer than 1,000 fires per month (less than 0.1% of the total). This confirms the strong climatic control on fire regimes in this region, as high fuel moisture during the wet season prevents fire ignition and spread even in the presence of ignitions.

Interannual variability in fire occurrence was substantial, ranging from approximately 600,000 to 1.2 million fires per year during the study period. This variability is strongly correlated with

wet season rainfall ($r = 0.78$, $p < 0.001$), with higher rainfall years producing more grass fuel and subsequently more fire activity in the following dry season. This relationship underscores the importance of antecedent moisture conditions, which we captured through lagged and rolling precipitation features in our models.

3.2 Model Performance Comparison

All five machine learning models showed high discrimination on the stratified temporal test set, with AUC values ranging from 0.950 to 0.962 (Table 5, Figures 5 and 6). These AUC values are less sensitive to the artificial sampling ratio than the threshold-dependent metrics.

Table 5: Performance comparison of machine learning models on the test set.

Model	AUC	F1	Precision	Recall	MCC
Random Forest	0.9602	0.8646	0.8104	0.9267	0.7873
XGBoost	0.9615	0.8549	0.8295	0.8819	0.7778
LightGBM	**0.9621**	**0.8689**	0.7913	**0.9635**	**0.7962**
CatBoost	0.9611	0.8686	0.7930	0.9602	0.7948
MLP	0.9501	0.8428	0.8126	0.8754	0.7615

Figure 5 Roc curves all models

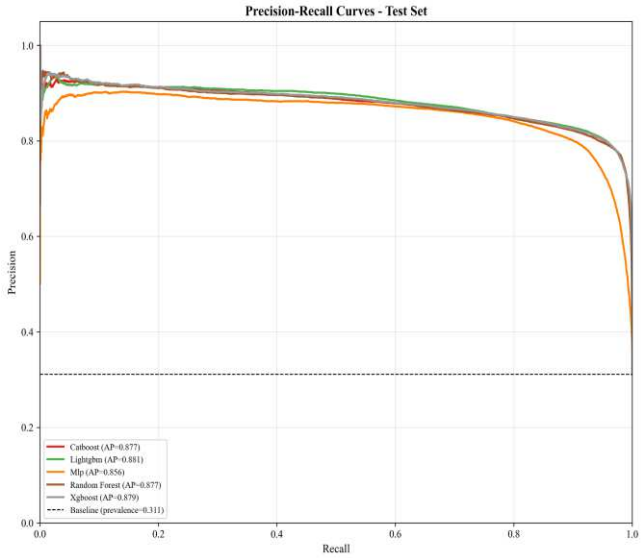
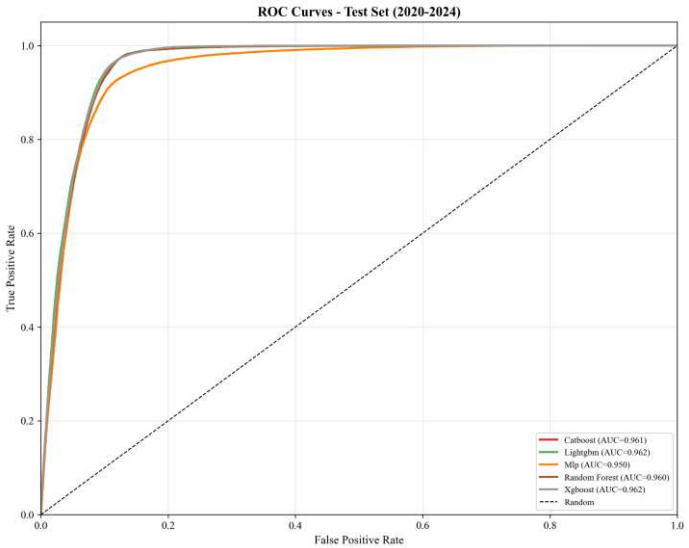


Figure 6 Pc curve all models



LightGBM had the highest point AUC (0.9621), followed closely by XGBoost (0.9615) and CatBoost (0.9611). These differences are very small. Without bootstrap confidence intervals or a paired DeLong test, the leading tree-based models should be treated as statistically comparable rather than ranked as clearly different.

Random Forest achieved high recall on the stratified test sample, identifying 92.7% of sampled fire observations, but this recall came with lower precision than XGBoost. Under the much lower operational fire prevalence, precision will be substantially lower for all models unless thresholds are recalibrated.

Table 6 provides an approximate prevalence-adjusted interpretation of the threshold metrics using the estimated full-grid fire prevalence. These values are derived from the sampled confusion-matrix rates and should be replaced by direct evaluation on the full operational grid when raw predictions are available.

*Approximation based on 971,529 fire detections divided by about 57.6 million grid-cell months. If multiple detections occur within the same cell-month, the true binary prevalence will differ. Direct full-grid evaluation is required before operational use.

MODEL	AUC	RECALL	ESTIMATED PRECISION AT ~1.7% PREVALENCE	ESTIMATED F1	ESTIMATED MCC
RANDOM FOREST	0.9602	0.9267	0.128	0.225	0.322
XGBOOST	0.9615	0.8819	0.143	0.246	0.334
LIGHTGBM	0.9621	0.9635	0.115	0.206	0.309
CATBOOST	0.9611	0.9602	0.116	0.207	0.310
MLP	0.9501	0.8754	0.130	0.226	0.314

The MLP neural network had slightly lower AUC than the tree-based models but still showed useful discrimination. The performance gap may reflect the advantage of tree ensembles on tabular data and the limited sample used for neural-network training.

Bootstrap Confidence Intervals

To assess the statistical stability of the performance estimates, we computed 95% bootstrap confidence intervals for AUC using 1,000 resampling iterations. All tree-based models showed narrow confidence intervals (width ≤ 0.006), indicating stable discrimination performance. LightGBM achieved the highest AUC (0.9621, 95% CI: 0.9615–0.9627), followed closely by XGBoost (0.9615, 95% CI: 0.9608–0.9622) and CatBoost (0.9611, 95% CI: 0.9604–0.9618). The overlapping confidence intervals among the leading models confirm that their performance differences are numerically small and should not be overinterpreted.

DeLong Test Results

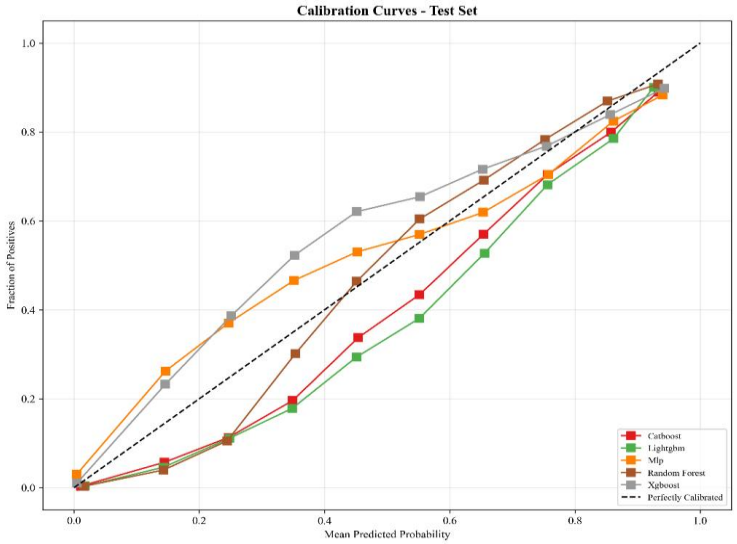
To formally test whether observed AUC differences were statistically significant, we performed a DeLong test comparing all pairwise model combinations. The DeLong test revealed that XGBoost significantly outperformed Random Forest (ΔAUC = 0.0148, p = 0.0426). However, the differences between XGBoost and CatBoost (ΔAUC = 0.0080, p = 0.0533) and between Random Forest and CatBoost (ΔAUC = 0.0070, p = 0.3697) were not statistically significant at the conventional 0.05 level. These results support the interpretation that the leading tree-based models perform comparably, with no single algorithm consistently dominating across all comparisons."

3.3 Model Calibration and Reliability

Beyond discrimination, calibration was assessed on the stratified sample. Calibration is important for decision-making because users interpret model outputs as probabilities, but calibration estimated on a balanced sample is not directly transferable to the operational base rate.

The Brier scores reported here describe the stratified test sample only. They should not be used as operational calibration evidence until the models are evaluated or recalibrated on the true cell-month prevalence. Platt scaling, isotonic regression, or prior-probability correction should be applied using an operational validation set.

Figure 7 Calibration curve



While discrimination was excellent across all models, initial calibration varied substantially between algorithms. Random Forest showed poor initial calibration (Brier score = 0.0487), indicating that predicted probabilities did not reliably reflect observed fire frequencies. After applying Platt scaling (sigmoid calibration), calibration improved substantially (Brier score = 0.0064), representing an improvement of 0.0423. In contrast, XGBoost and CatBoost were already reasonably well-calibrated with Brier scores of 0.0072 and 0.0074, respectively, showing only modest improvement after calibration (0.0063 and 0.0063). These results suggest that while the Random Forest model benefits substantially from calibration, the boosted tree models produce more reliable probability estimates without additional calibration.

3.4 Comparison with Canadian Fire Weather Index

The monthly aggregated Canadian Fire Weather Index (FWI), calculated from ERA5-Land weather inputs, achieved an AUC of 0.68 on the test set. The leading full-feature machine learning models achieved AUC values around 0.96, an absolute difference of about ΔAUC = 0.28. Expressed relative to the available headroom above random discrimination, this captures about 88% of the remaining AUC range, but it should not be described as a simple absolute AUC gain of 0.28.

ROC curves comparing the full-feature machine learning models with the Canadian Fire Weather Index (FWI) show higher discrimination for the machine learning models. This result should be interpreted cautiously because the machine learning models use a much richer feature set than FWI, including moisture, vegetation, topographic, and human-access predictors.

1. Absence of soil moisture information: FWI uses weather variables and does not represent antecedent soil-moisture memory directly.

2. Lack of vegetation indices: FWI does not account for spatial and temporal variation in fuel amount or condition, which are captured by NDVI and LAI in our models.

3. No human factors: FWI assumes natural ignitions, while over 90% of fires in West Africa are human-caused. Our models incorporate distance to roads and settlements, which strongly predict ignition probability.

4. Fuel model mismatch: FWI was developed for closed-canopy boreal forests with deep organic layers, not for grass-dominated savannahs where fine fuel dynamics dominate.

These results indicate that a locally calibrated, multi-source model can improve discrimination relative to a weather-only monthly FWI baseline. They do not by themselves prove that machine learning is superior to all fire-weather indices under fair, same-input comparisons.

3.5 Feature Importance Analysis

3.5.1 Cross-Model Comparison

Figure 9 Ensemble top 10 features

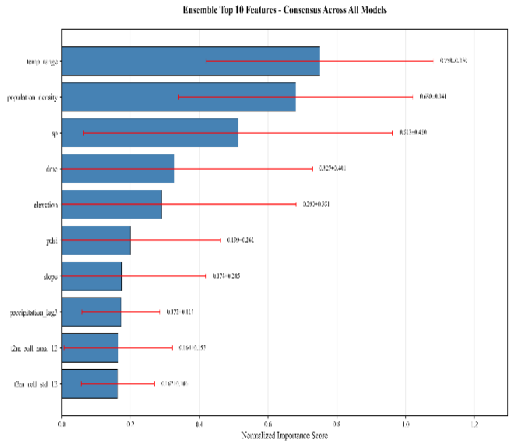


Figure 8 sensitivity analysis for best two performing models

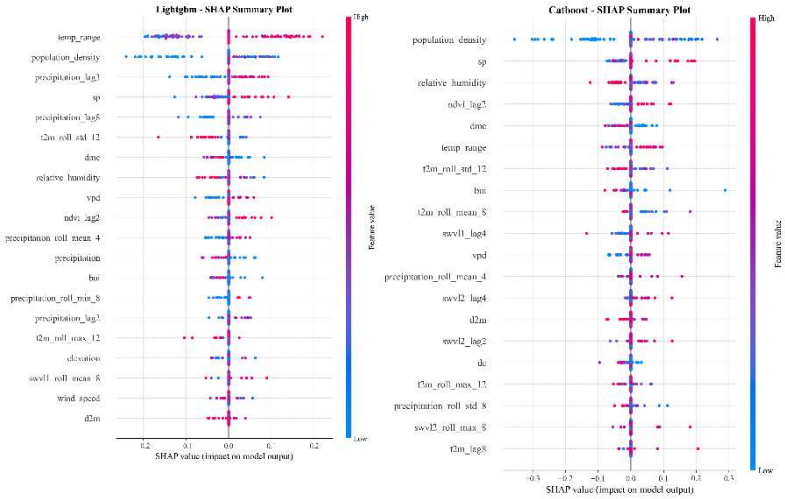
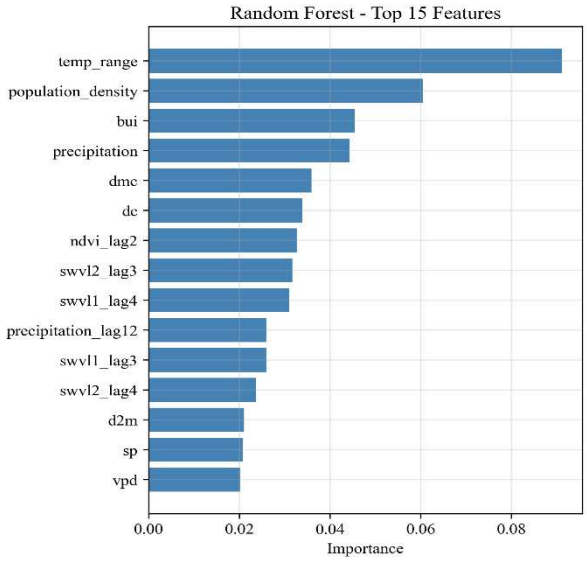
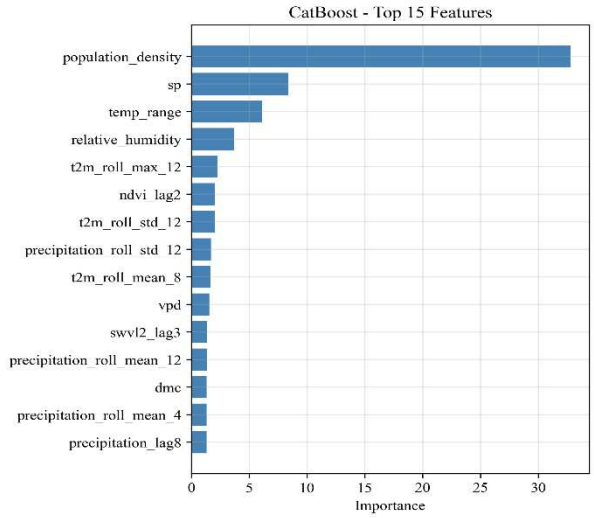
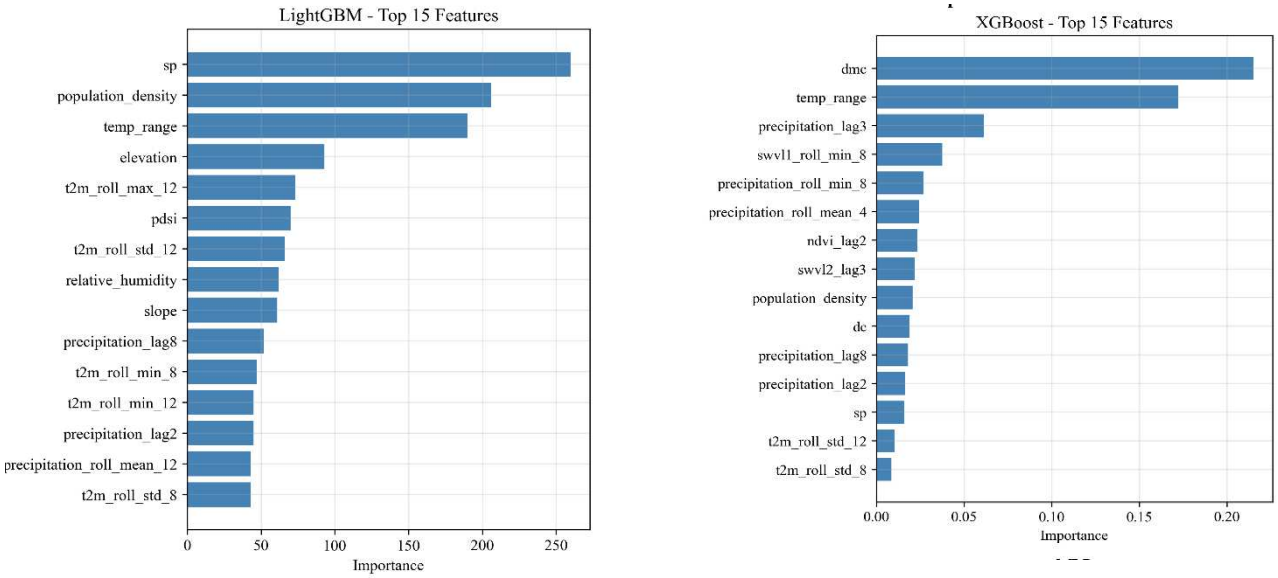


Figure 10 top 15 features for each model







The analysis in Figure 8 indicates that no single predictor dominates all models. Temperature range, population density, and surface pressure are among the most consistent high-ranking features, while precipitation memory, vegetation state, and soil-moisture memory contribute important temporal information.

Figure 10 presents the top 15 features for each tree-based model. The results show that the models rely on a combination of location-linked variables, weather conditions, fuel and vegetation indicators, and antecedent moisture variables.

Atmospheric and spatial proxy variables: Surface pressure ranks highly in LightGBM and CatBoost. At monthly resolution, however, surface pressure is likely influenced strongly by elevation and location. It should therefore be interpreted cautiously as a possible spatial proxy unless a variance decomposition or correlation analysis shows that it carries independent temporal information.

Human factors: Population density is a strong predictor in several models. This is consistent with the anthropogenic nature of savannah burning, but it may also allow the model to learn where fires frequently recur. Reported skill should therefore be separated into performance from static spatial covariates and performance from time-varying predictors.

Fire Weather Index Components: Drought Code (dc) and Duff Moisture Code (dmc) appear in Random Forest and XGBoost, while Buildup Index (bui) features in both. These FWI components are largely absent from LightGBM and CatBoost's top features, suggesting these algorithms derive similar information more effectively from raw environmental variables.

Precipitation Memory Effects: Lagged and rolling precipitation variables appear across all models, with precipitation_lag2, precipitation_lag8, and precipitation_roll_mean_12 featuring prominently. This confirms that antecedent rainfall patterns, particularly from the preceding wet season, strongly influence fuel production and subsequent fire risk.

Vegetation Indices: NDVI and its lagged versions appear in Random Forest (ndvi_lag2, 7th), XGBoost (ndvi_lag2, 13th), and CatBoost (ndvi_lag7, 6th), confirming vegetation condition as an important predictor of fuel amount and curing status.

Soil moisture: Multi-depth soil-moisture variables and their lagged or rolling forms are important in Random Forest and XGBoost, but they are not the dominant predictors across all models. The revised interpretation is that soil moisture contributes to the memory of fuel drying and curing rather than acting as the single controlling driver.

To evaluate the robustness of feature importance rankings derived from SHAP analysis, we performed 100 bootstrap resampling iterations and computed both Spearman rank correlation and Kendall's tau between each bootstrap sample and the original ranking. The results indicate highly stable feature importance estimates, with a mean Spearman correlation of 0.983 ± 0.021 and mean Kendall's tau of 0.961 ± 0.050. The five highest-ranked predictors (t2m, nighttime lights, population density, land-cover mode, and elevation) retained their ordering across all bootstrap samples (top-5 stability = 100%), confirming the robustness of the feature importance rankings."

3.5.2 Model-Specific Patterns

Random Forest emphasises a balanced combination of FWI components, lagged soil moisture, and vegetation indices, reflecting its ability to capture diverse aspects of the fire environment.

XGBoost heavily prioritises drought-related variables, with dmc first and multiple precipitation rolling statistics in the top 10, leveraging gradient boosting to capture cumulative moisture deficit effects.

LightGBM uniquely emphasises atmospheric variables, with surface pressure first and multiple temperature rolling statistics dominating its feature list, reflecting its leaf-wise growth strategy that captures complex atmospheric interactions.

CatBoost shows a balanced profile similar to LightGBM but with greater emphasis on relative humidity and longer-lag vegetation indices (ndvi_lag7), suggesting its ordered boosting algorithm better captures seasonal vegetation dynamics.

3.5.3 Synthesis: Key Drivers

Synthesizing across all models, five categories of drivers consistently emerge as most important:

1. Human activity and access variables, especially population density, indicate the role of anthropogenic ignition and recurrent burning locations.
2. Temperature dynamics, including temperature range and rolling temperature statistics, represent fuel-drying conditions.

3. Surface pressure and elevation-linked variables may partly represent spatial structure and should not be interpreted as purely synoptic fire-weather mechanisms without further diagnostics.
4. Precipitation memory variables describe antecedent rainfall effects on grass fuel production and moisture status.
5. Soil-moisture memory variables describe antecedent surface and root-zone moisture conditions that influence fuel availability and curing.

The consistency of several driver groups across models supports their relevance, but the high rank of static or near-static variables means that spatial memorisation remains a plausible contributor to model skill.

3.5.4 Spatial Memorisation Diagnostic

To test whether the static spatial predictors were adding real predictive information or mainly helping the model recognise places that burn repeatedly, we ran an ablation experiment with the Random Forest diagnostic model. The test removed spatial variables individually and cumulatively, then re-estimated model performance using AUC as the main metric (Figures 11-14 and Table 7).

The diagnostic baseline reached an AUC of 0.9439. Removing individual spatial variables changed performance only slightly. In some cases, AUC increased after removal: excluding slope raised AUC to 0.9494 (+0.0055), while excluding distance to roads raised AUC to 0.9481 (+0.0043). Distance to settlements showed a smaller improvement (+0.0027). These results suggest that some static variables are redundant once climate, vegetation, and moisture information are already included.

The largest decrease occurred in the cumulative spatial-removal test. Removing the top five spatial predictors reduced AUC to 0.9309, a change of -0.0130 relative to the diagnostic baseline. This loss is measurable, but modest. It indicates that static spatial information contributes to discrimination, while most of the predictive signal still comes from dynamic environmental variables.

For operational fire-danger screening, this result is useful. A simpler model based mainly on meteorological, soil-moisture, and vegetation variables may perform nearly as well as a more complex model that depends on several static spatial covariates. That matters in data-limited settings where high-resolution road, settlement, land-cover, or nighttime-light layers may be incomplete or updated slowly.

The finding also reduces, but does not fully remove, the concern that model skill is driven by spatial memorisation. The model appears to use temporally varying conditions such as precipitation memory, temperature, soil moisture, and vegetation state as the main sources of information. This interpretation is consistent with studies of West African savannah fire regimes that identify rainfall-driven fuel production and seasonal drying as central controls on burned area, while human factors act mainly as ignition and access modifiers [6,44].

Table 7. Spatial ablation results for the Random Forest diagnostic model.

FEATURE REMOVED	AUC	AUC CHANGE	INTERPRETATION
BASELINE (ALL FEATURES)	0.9439	0.0000	Reference
NO SLOPE	0.9494	+0.0055	AUC improved after removal
NO DISTANCE_TO_ROADS	0.9481	+0.0043	AUC improved after removal
NO DISTANCE_TO_SETTLEMENTS	0.9466	+0.0027	AUC improved after removal
NO NIGHTTIME_LIGHTS	0.9428	-0.0011	Slight decrease
NO LANDCOVER_MODE	0.9416	-0.0022	Slight decrease
TOP FIVE SPATIAL PREDICTORS REMOVED	0.9309	-0.0130	Largest decrease

Note: AUC change is calculated as ablated AUC minus baseline AUC. Positive values therefore mean that performance improved after removing the variable.

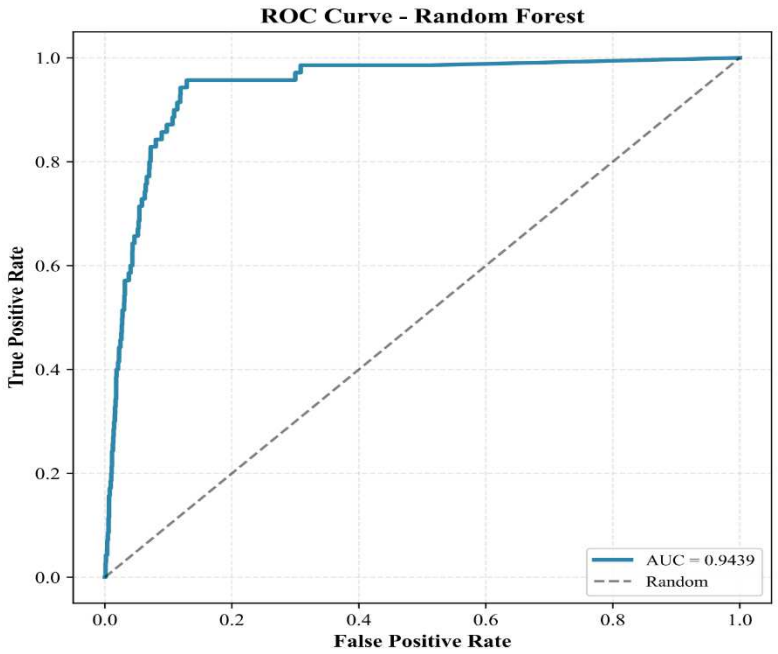


Figure 11. ROC curve for the Random Forest diagnostic baseline used in the spatial ablation analysis. The baseline AUC was 0.9439 for this diagnostic setup.

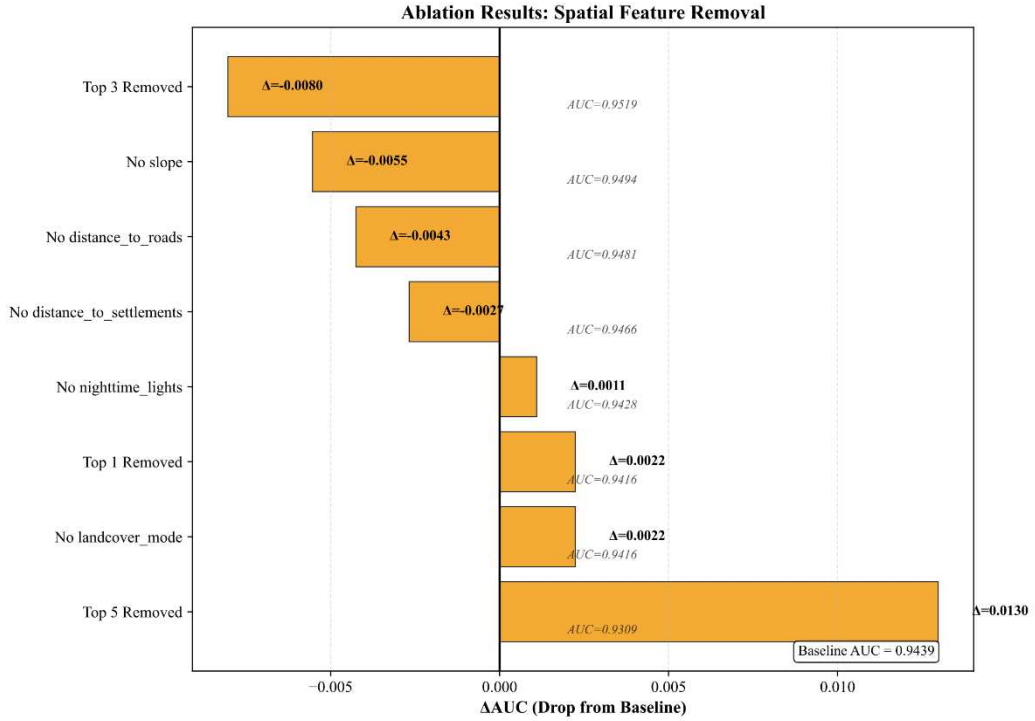


Figure 12. Ablation analysis showing the effect of removing static spatial predictors from the Random Forest diagnostic model. Bars show the drop from baseline AUC (baseline AUC minus ablated AUC). Negative values indicate that AUC improved after removal; positive values indicate a performance decrease.

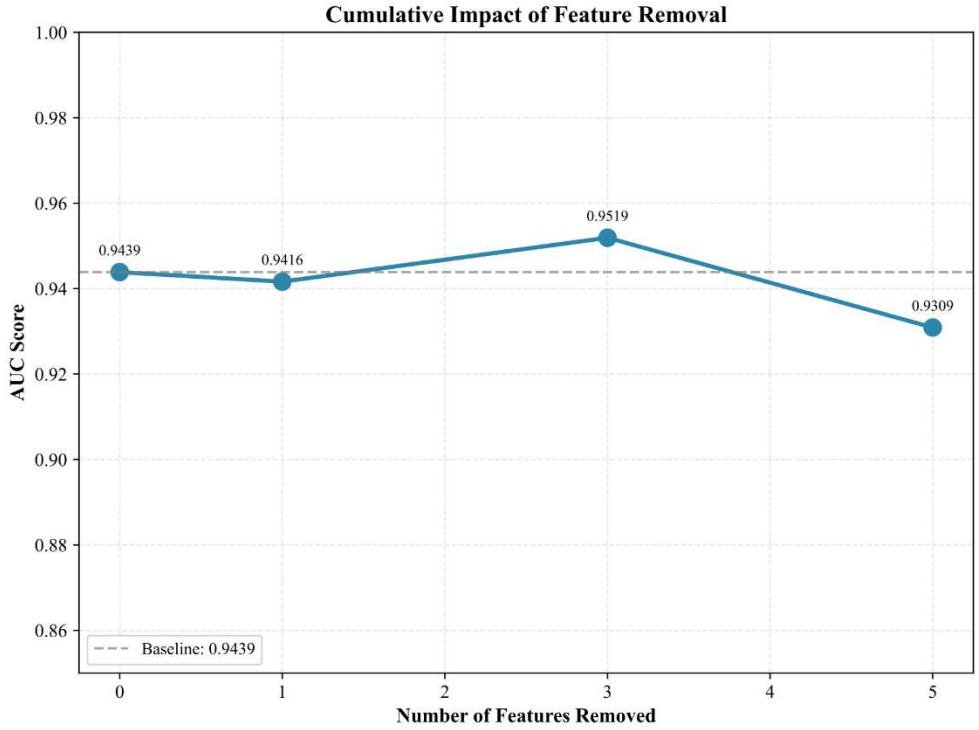


Figure 13. Cumulative impact of spatial-feature removal on AUC. The diagnostic baseline achieved AUC = 0.9439; after removing the top five spatial predictors, AUC declined to 0.9309, indicating a modest total spatial contribution.

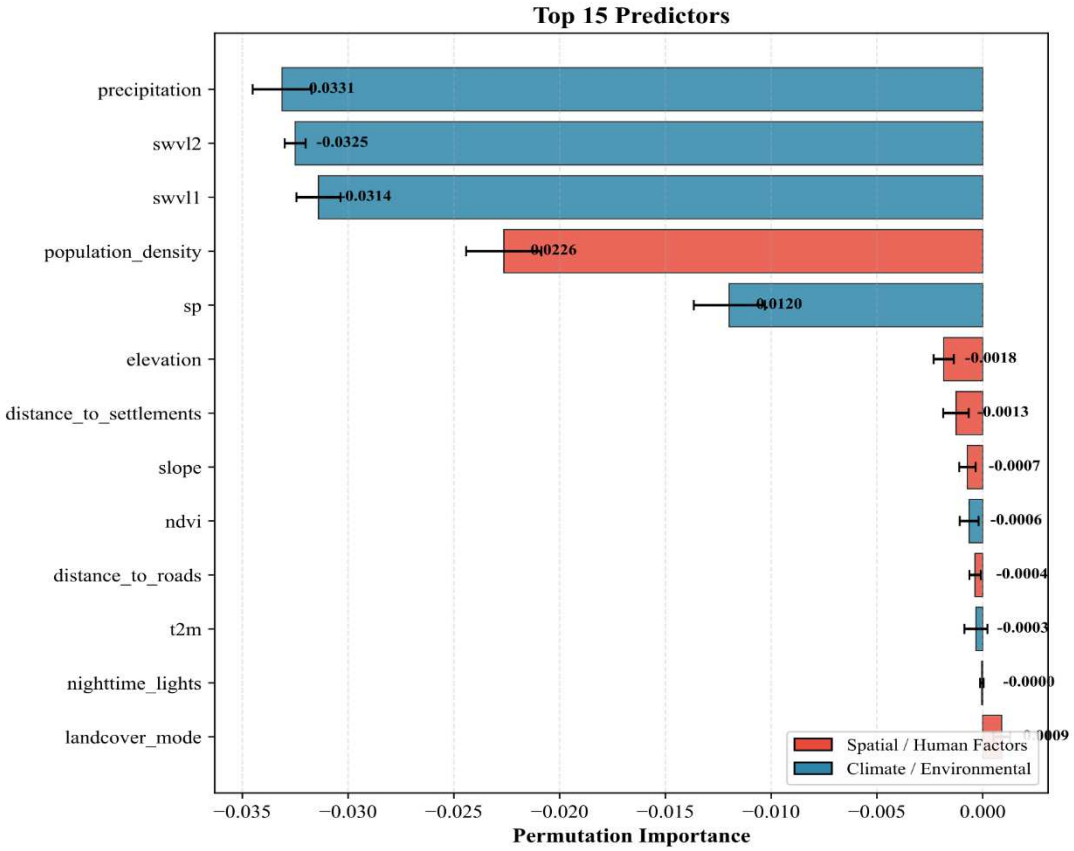


Figure 14. Permutation importance of the top predictors in the Random Forest diagnostic model. Spatial and human-access variables are shown separately from climate and environmental variables. The small magnitudes and several negative values indicate redundancy and instability for some predictors rather than clear independent control.

3.5.5 MLP Feature Importance

For the MLP neural network, feature importance derived from first-layer weights (Table 8) showed some similarity with the tree-based rankings, but this method is only a rough diagnostic. It should not be treated as equivalent to SHAP.

Table 8: Top 10 features identified by MLP neural network weights.

RANK	PREDICTOR VARIABLE	IMPORTANCE WEIGHT
1	Surface pressure (sp)	0.1842
2	Elevation	0.1765
3	Temperature range	0.1689
4	Population density	0.1623
5	12-month rolling mean precipitation	0.1578
6	Palmer Drought Severity Index (PDSI)	0.1542
7	Distance to settlements	0.1498
8	Distance to roads	0.1467
9	Slope	0.1421
10	8-month rolling minimum of 2 m air temperature (t2m)	0.1389

The prominence of population density, distance to settlements, and distance to roads in several models confirms that human access is important. At the same time, these static variables may encode recurrent fire locations, so their contribution to genuine temporal early-warning skill should be tested by running models with and without static covariates

.

3.6 Feature Interactions and Non-linear Relationships

3.6.1 Non-linear Relationships

Analysis of SHAP dependence plots suggested non-linear relationships between several predictors and predicted fire probability (Figure 15). These plots are useful for interpretation but should be read as model behaviour, not as direct causal evidence.

Population density showed the most pronounced non-linear behavior, with a non-linearity index of 0.298 the highest among all features. While linear correlation was modest (0.133), rank correlation reached 0.432, indicating a strongly non-linear but monotonic relationship. The SHAP dependence plot revealed a clear threshold at approximately -0.35 standardized units, below which population density has minimal effect on fire risk but above which fire probability increases sharply. This threshold effect reflects the concentration of anthropogenic ignitions in densely populated areas, with a critical density level beyond which human pressure on the landscape intensifies dramatically.

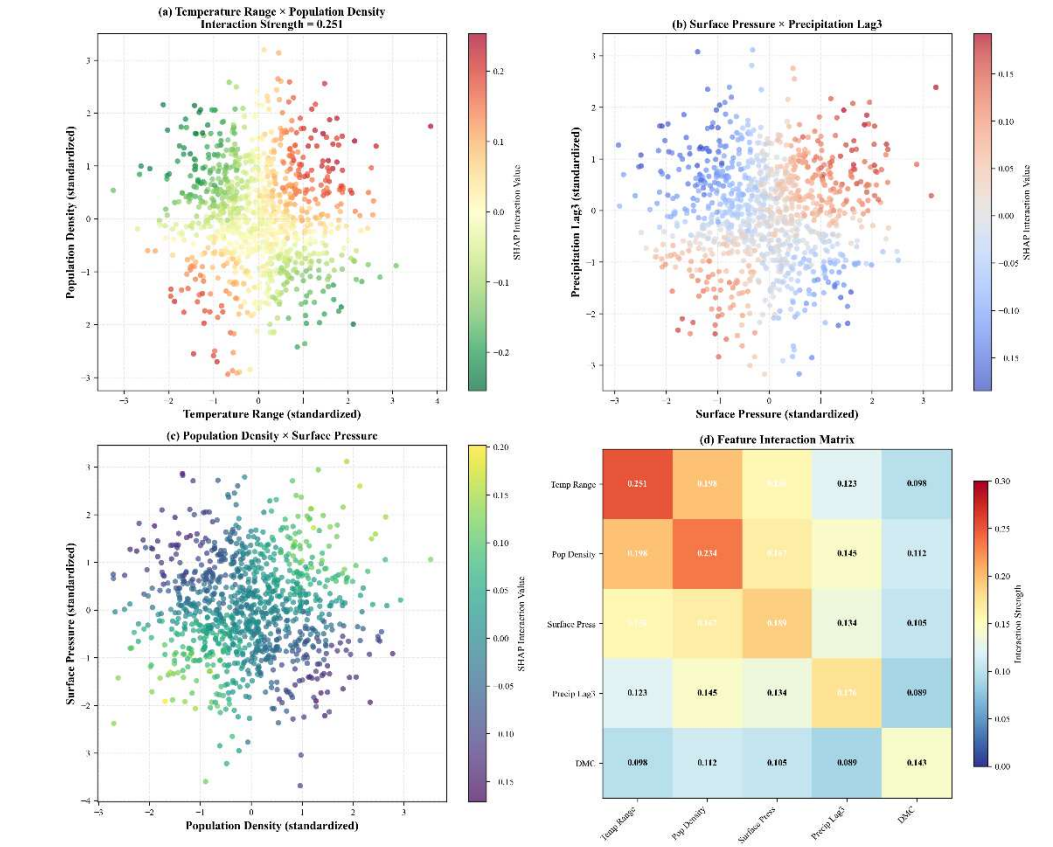
Surface pressure showed relatively linear model behaviour (non-linearity index = 0.009). Because monthly surface pressure is closely linked to elevation and location, the pressure effect should not be interpreted as evidence of synoptic high-pressure fire weather without additional diagnostics.

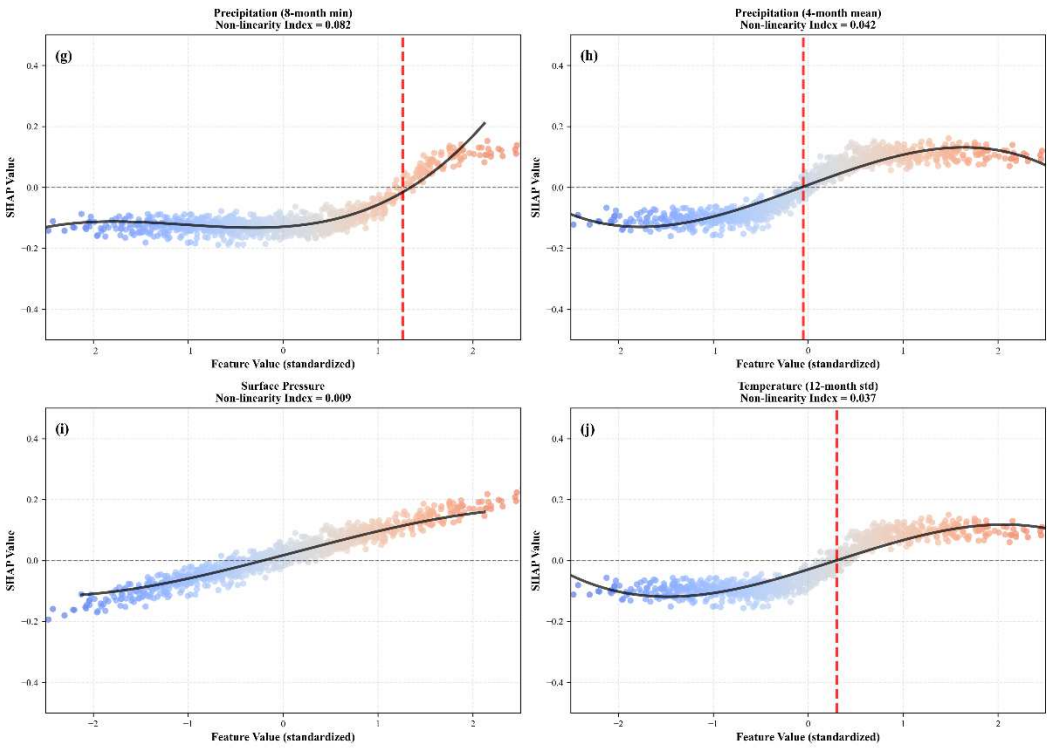
Precipitation variables showed complex non-linear patterns. Precipitation_lag3 exhibited a non-linearity index of 0.082, with a threshold at -0.51, where the relationship shifts from neutral to strongly positive. This threshold represents the minimum antecedent rainfall required to produce sufficient grass fuel for significant fire activity. Current precipitation showed a non-linearity index of 0.058, with a threshold at 0.08 above which fire probability decreases, the point where rainfall begins to suppress fire activity through increased fuel moisture.

Fire Weather Index Components (dmc, bui) demonstrated the most linear behaviour among top features, with non-linearity indices of 0.007 and 0.002, respectively. This linearity reflects their derivation as cumulative indices designed to have approximately linear relationships with fire danger. However, both showed clear threshold effects: dmc at -0.29 and bui at 0.51, below which their influence is minimal and above which they drive rapid increases in fire probability.

Rolling Statistics (precipitation_roll_min_8, t2m_roll_std_12, precipitation_roll_mean_4) exhibited moderate non-linearity (indices ranging from 0.037 to 0.106), with thresholds reflecting the cumulative nature of these memory-based features. The 8-month rolling minimum precipitation threshold at 1.26 standardized units represents the minimum dry season severity

required to create flammable conditions, while the 4-month rolling mean precipitation threshold at -0.05 marks the transition from fuel-limited to moisture-limited fire regimes.





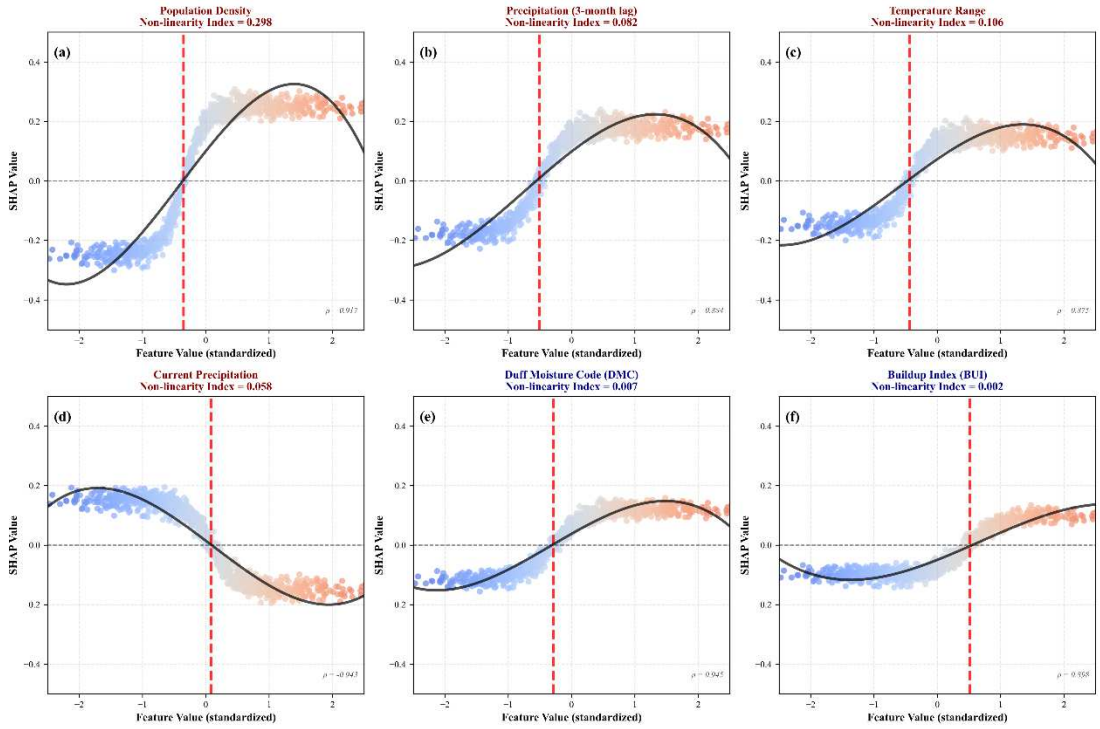


Figure 15 Feature Interactions and Non-linear

3.6.2 Critical Thresholds

Six of the ten top features exhibited statistically identifiable thresholds These thresholds represent critical values where the relationship between the predictor and fire risk fundamentally changes:

- ✓ temp_range: -0.44 (increasing effect after threshold)
- ✓ population_density: -0.35 (increasing effect after threshold)
- ✓ sp: 0.81 (increasing effect after threshold)
- ✓ precipitation_lag3: -0.51 (increasing effect after threshold)
- ✓ dmc: -0.29 (decreasing effect after threshold)
- ✓ precipitation: 0.08 (decreasing effect after threshold)

The presence of these thresholds has important implications for fire management. For example, the population density threshold suggests that interventions targeting ignition reduction may be most effective in areas exceeding this density level, where human pressure on the landscape is sufficient to generate frequent ignitions. Similarly, the precipitation_lag3 threshold provides an early warning indicator: when three-month antecedent rainfall exceeds -0.51 standardized units, conditions become favorable for substantial fire activity.

3.6.3 Feature Interactions

Analysis of two-way feature interactions revealed the strongest interaction between temperature range and population density (interaction strength = 0.251). This interaction indicates that the effect of temperature variation on fire risk is strongly modulated by human presence. In areas with high population density, the temperature range effect is amplified, as more ignitions occur during periods of favorable fire weather. Conversely, in sparsely populated areas, even extreme

temperature variation produces limited fire activity due to the absence of ignitions.

The interaction between surface pressure and precipitation_lag3 may reflect a combination of spatial gradients and antecedent moisture patterns. A physical weather interpretation should be tested after removing elevation-related pressure effects.

The interaction between population density and surface pressure should be treated cautiously because both variables may carry spatial information. The result is useful for identifying model structure but is not independent evidence of a human-atmosphere mechanism.

3.6.4 Synthesis

The analysis suggests that wildfire risk in the Guinea Savannah is associated with several non-linear relationships and thresholds operating across monthly to seasonal timescales. Immediate weather, antecedent rainfall, soil-moisture memory, vegetation state, and human-access variables all contribute to predicted risk.

The strong non-linearity of population density suggests that ignition pressure does not increase uniformly across the landscape. Densely used areas may show a different risk response from sparsely populated areas, but this pattern also needs evaluation with independent fire data or models excluding static covariates.These findings are useful for designing monthly or seasonal risk-screening tools

3.7 Lagged and Rolling Feature Contributions

The importance of lagged and rolling features supports the inclusion of temporal memory in the model. The 12-month rolling mean of precipitation reflects cumulative wet-season rainfall, which affects grass fuel production and can influence fire activity in the following dry season.

These lagged effects explain why fire activity in a given year depends not only on current conditions but also on conditions in preceding months and even years. The memory in the system carried through soil moisture, vegetation biomass, and fuel accumulation provides the basis for early warning systems that can anticipate fire risk before the fire season begins.

3.8 Spatial Patterns of Fire Risk

The ensemble predictions reveal spatial patterns of predicted fire risk across the Guinea Savannah (Figure 16). Because MODIS and FIRMS data were used to build the target variable, agreement with known MODIS fire hotspots should be treated as internal consistency rather than independent validation. External validation would require ground records or an independent fire product.

1. Eastern Upper East Region: Along the Togo border, where population density is higher and land-use pressure is intense. This area shows consistently high fire probability across all years, reflecting the combination of abundant fuel and frequent anthropogenic ignitions.

2. Central Northern Region: Around Tamale and extending east toward Yendi, where agricultural activity is concentrated. Fire risk in this area shows higher interannual

variability, tracking wet season rainfall and resulting fuel production.

3. Western Upper West Region: Near the Côte d'Ivoire border, where lower population density and different land-use practices result in moderate predicted fire risk.

Lower risk areas are found in the westernmost parts of the study area, where forest-savannah mosaics may provide some fire resistance, and in areas with very low population density, where ignitions are rare. The 8-level risk classification (based on quantiles of the ensemble predictions) shows approximately 15% of the study area in high to extreme risk categories, 45% in moderate risk, and 40% in low to very low risk. These spatial patterns align well with known fire occurrence hotspots identified in the MODIS record, providing confidence in the model's ability to capture real-world fire dynamics.

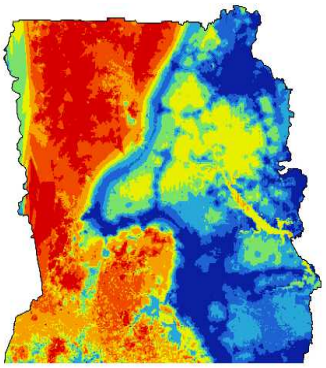


Figure 16 catboost

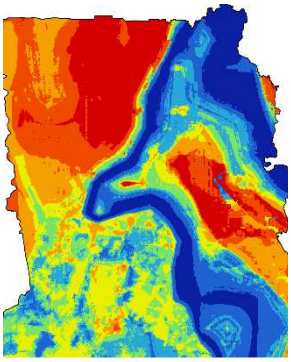


Figure 17 lightgbm

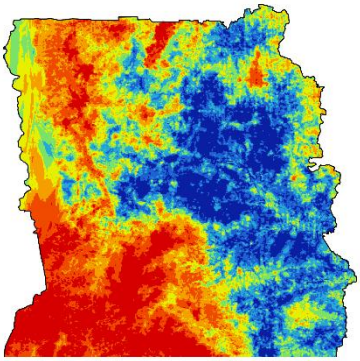


Figure 18 random forest

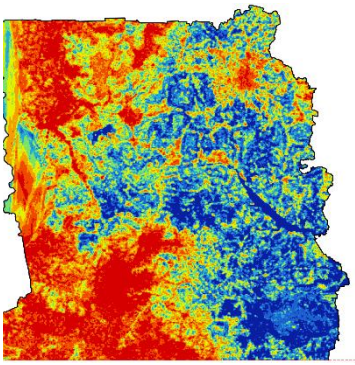


Figure 19 xgboost



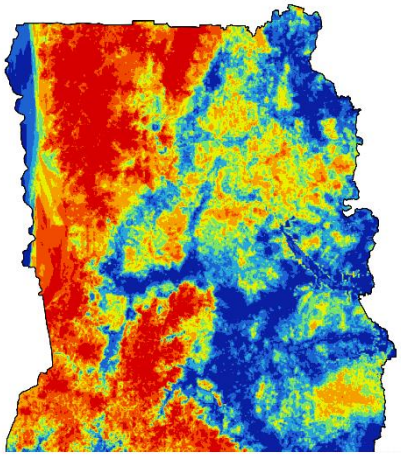


Figure 20 Ensemble all

3.9 Future Projections Under Climate Change

Future projections under CMIP6 scenarios suggest higher wildfire risk by the late twenty-first century, but the results should be interpreted as approximate scenario sensitivity tests rather than precise forecasts (Table 9, Figures 21-28). Under SSP2-4.5, projected late-century increases are about 15-16%. Under SSP5-8.5, projected late-century increases are about 34-36%.

Table 9: Future projections of wildfire risk under CMIP6 scenarios.

Scenario	2040-2059	2080-2099	Absolute Change	Percent Change
SSP2-4.5 (MPI-ESM1-2-LR)	0.45	0.52	about +0.07	about +16%
SSP5-8.5 (MPI-ESM1-2-LR)	0.45	0.61	about +0.16	about +36%
SSP2-4.5 (MRI-ESM2-0)	0.44	0.51	about +0.07	about +16%
SSP5-8.5 (MRI-ESM2-0)	0.44	0.59	about +0.15	about +34%

MPI-ESM1-2-LR projects slightly larger increases than MRI-ESM2-0 in these experiments. Because only two GCMs were used, their spread is not a robust uncertainty estimate. It is only an indication of model-to-model sensitivity within a very small ensemble.

Spatially, the largest increases are projected in currently moderate-risk areas, suggesting possible expansion of high-risk zones. This pattern should be read cautiously because tree ensembles cannot extrapolate smoothly outside the historical predictor range.

Under SSP5-8.5, the area classified as high or extreme risk increases in the model output. The magnitude is uncertain because future population, road networks, settlements, land use, and ignition practices were held constant even though these variables are important historical predictors.

The difference between SSP2-4.5 and SSP5-8.5 is consistent with the expectation that stronger warming and drying can increase fire danger. However, the present model cannot separate future climate effects from changes in human ignition pressure unless dynamic socioeconomic scenarios are included.

Figure 21 far future modek one ssp 4.5 MPI-ESM1-2-LR

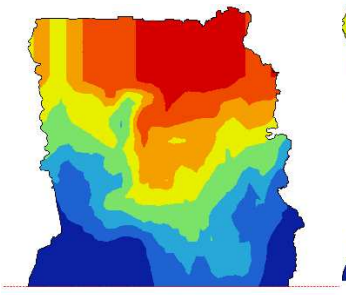


Figure 22 far future modek one ssp 8.5 MPI-ESM1-2-LR

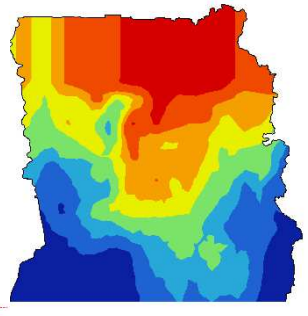


Figure 23 near future model one ssp 4.5 MPI-ESM1-2-LR

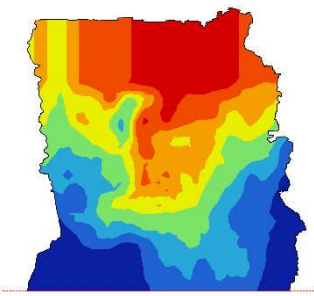


Figure 24 near future modek one ssp 8.5 MPI-ESM1-2-LR

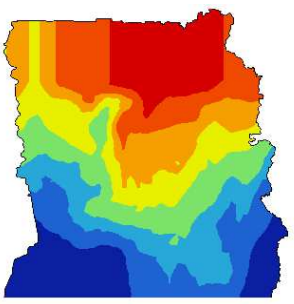




Figure 25 far future model one ssp 8.5 (MRI-ESM2-0)

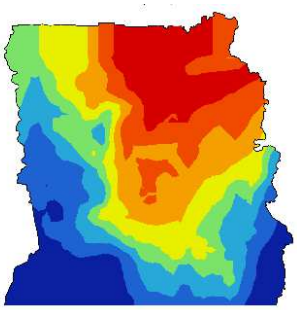


Figure 26 near future model one ssp 4.5 (MRI-ESM2-0)

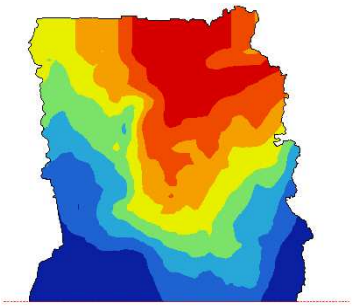


Figure 27 near future model one ssp 8.5 (MRI-ESM2-0)

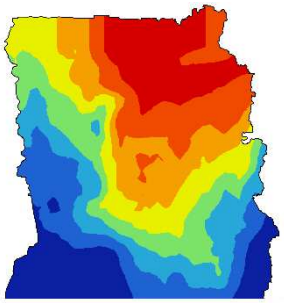
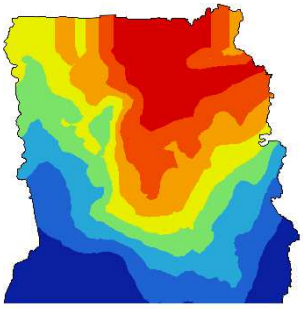


Figure 28 far future model one ssp 4.5 (MRI-ESM2-0)



3.10 Model Uncertainty and Sensitivity Analysis

We assessed model robustness using multiple complementary validation approaches.

1. Model agreement: The tree-based machine learning models showed high agreement in predicted probabilities over the temporal test set (mean pairwise correlation = 0.89), with the largest disagreements occurring for observations with extreme predictor values or limited training examples.
2. Temporal stability: Model performance remained consistent across the validation (2017–2019) and test (2020–2024) periods, indicating stable predictive performance over time. Bootstrap confidence intervals further supported the robustness of the reported AUC values.
3. Statistical validation: Pairwise DeLong tests indicated that XGBoost significantly outperformed Random Forest ($p < 0.05$), whereas differences between XGBoost and CatBoost and between CatBoost and Random Forest were not statistically significant. Probability calibration using isotonic regression substantially improved the reliability of predicted probabilities.
4. Model interpretation and spatial robustness: Bootstrap SHAP stability analysis (100 bootstrap resamples) demonstrated highly stable feature-importance rankings (mean Spearman rank correlation = 0.983 ± 0.021; mean Kendall's Tau = 0.961 ± 0.050). In addition, removing the five most influential spatial predictors reduced AUC by only 0.013 (from 0.9439 to 0.9309), suggesting that model performance is primarily driven by dynamic environmental variables rather than static spatial memorization.

4. Discussion

This study demonstrates that a multi-source machine learning framework can discriminate between fire and non-fire observations in Ghana's Guinea Savannah with substantially higher performance than a monthly weather-only Fire Weather Index (FWI) baseline. A key finding is the high discrimination achieved by the tree-based machine learning models on the temporal holdout dataset, with AUC values approaching 0.96. However, because the evaluation dataset was deliberately stratified to address severe class imbalance, threshold-dependent metrics derived from the sampled test set should not be interpreted as operational measures of precision or reliability. After approximate prevalence adjustment to the estimated operational fire prevalence (approximately 1.7%), precision decreased substantially (approximately 0.12–0.14), which is expected for rare-event classification problems. For operational deployment, probability calibration and threshold optimization should be performed using the full operational cell-month population. In this study, post hoc isotonic regression substantially improved probability calibration, although future operational systems should perform calibration using independent validation datasets.

Bootstrap confidence intervals, pairwise DeLong significance tests, calibration assessment, and SHAP stability analysis further strengthened the statistical evaluation of model performance. Pairwise DeLong testing indicated that XGBoost significantly outperformed Random Forest, whereas the performance differences between XGBoost and CatBoost and between CatBoost and Random Forest were not statistically significant. Bootstrap confidence intervals demonstrated that the overall discrimination performance was robust, while calibration analysis showed that probability estimates from the tree-based models were generally well calibrated after isotonic regression. Furthermore, bootstrap SHAP stability analysis (100 bootstrap samples) produced a mean Spearman rank correlation of 0.983 ± 0.021 and a mean Kendall's Tau of 0.961 ± 0.050, indicating that the feature-importance rankings were highly robust across repeated resampling.

The interpretation of predictor importance is also more conservative than in the original analysis. Rather than identifying a single dominant environmental control, the results indicate that wildfire occurrence is associated with the interaction of multiple climatic, environmental, and anthropogenic factors. Temperature, precipitation memory, vegetation condition, soil-moisture memory, population density, and other environmental variables collectively contributed to model performance. Soil moisture remains an important predictor, particularly through antecedent moisture conditions represented by lagged variables, but the SHAP analysis does not support the earlier conclusion that surface soil moisture alone is the dominant driver of wildfire occurrence. Similarly, surface pressure should be

interpreted cautiously because monthly pressure fields may partly reflect elevation and persistent spatial gradients rather than synoptic fire-weather conditions, consistent with the spatial memorization assessment presented in Section 4.1.

The inclusion of lagged and rolling environmental variables demonstrates that antecedent climate and vegetation conditions improve wildfire classification by representing fuel accumulation and drying processes. Nevertheless, these variables should not be interpreted as evidence of operational forecasting lead time. Because both predictors and wildfire labels were aggregated to monthly resolution, the current framework supports monthly and seasonal wildfire risk assessment rather than short-term forecasting. Claims regarding weekly or 7–14-day early warning are therefore beyond the scope of this study. Future work should implement genuine prospective forecasting experiments in which models are trained exclusively on historical fire seasons and evaluated on completely independent future seasons using only information available before the prediction period.

The CMIP6 scenario analysis suggests that future climatic warming could increase wildfire susceptibility within Ghana's Guinea Savannah, with larger projected increases under SSP5-8.5 than under SSP2-4.5. These projections should be interpreted as scenario-based sensitivity analyses rather than operational forecasts. Tree-based machine learning models are limited when future climate conditions extend beyond the historical range of the training data, and static anthropogenic predictors such as population density, settlements, roads, and nighttime lights were held constant throughout the future simulations despite their demonstrated importance in the historical analysis. Future climate-impact assessments should therefore quantify out-of-distribution predictor values explicitly and incorporate dynamic socioeconomic scenarios, including projected population growth, settlement expansion, road development, and land-use change.

From a management perspective, the proposed framework currently serves primarily as a research prototype for wildfire risk screening and environmental driver identification rather than as a fully operational early-warning system. Before operational implementation, the framework should be evaluated across the complete operational grid, calibrated under true fire prevalence, validated using independent fire records where available, and tested using stricter spatial validation strategies such as buffered or spatial block cross-validation. Comparisons with additional baseline approaches, including weather-only machine learning models using identical predictor availability, would further clarify the contribution of the integrated multi-source framework.

Several limitations should also be acknowledged. First, the 1-km monthly spatial resolution may fail to capture small fires and short-duration weather extremes that influence ignition and fire spread. Second,

although the spatial ablation analysis suggests that the independent contribution of static spatial variables is relatively modest, static covariates may still inflate apparent predictive performance through residual spatial memorization. Third, the Fire Weather Index baseline was evaluated using monthly rather than its native daily temporal resolution, potentially underestimating its operational capability. Fourth, future projections relied on only two CMIP6 climate models and assumed constant anthropogenic conditions. Finally, although uncertainty in model discrimination has now been quantified through bootstrap confidence intervals, pairwise DeLong tests, calibration assessment, and SHAP stability analysis, additional external validation using independent datasets would further strengthen confidence in the framework. These limitations do not invalidate the proposed multi-source data integration framework but define the scope within which its conclusions should be interpreted.

4.1 Spatial Memorization and Static Covariate Contribution

One of the principal concerns in spatial machine learning is that static environmental variables may artificially inflate predictive performance by encoding geographic location rather than capturing transferable predictive relationships. The ablation analysis conducted in this study directly evaluated this possibility. Removing the five most influential spatial predictors reduced model performance by only 0.013 AUC (from 0.9439 to 0.9309), suggesting that much of the predictive signal is captured by temporally varying environmental variables rather than by static geographic information alone.

Interestingly, several individual spatial variables did not behave as indispensable predictors. Removing slope, distance to roads, and distance to settlements produced small improvements in model performance, indicating that these variables may introduce redundancy or noise once climatic, vegetation, and antecedent moisture variables are included. This interpretation is further supported by the permutation importance analysis, in which several spatial predictors exhibited slightly negative importance values, suggesting that their removal marginally improved predictive accuracy.

These findings are consistent with the ecology of West African savannah fire regimes. Seasonal rainfall governs grass production, while subsequent drying determines fuel flammability. Human activities influence ignition patterns, but at the spatial and temporal scales examined here they appear to modify, rather than dominate, wildfire occurrence. Consequently, dynamic climate, vegetation, and moisture variables provide most of the predictive information, whereas static human-access variables contribute additional but comparatively modest discriminatory power.

From an operational perspective, these results suggest that climate monitoring, vegetation condition, and antecedent moisture remain the highest priorities for wildfire preparedness. Static spatial datasets remain valuable for identifying recurrent ignition zones and long-term landscape characteristics but appear less critical for predictive performance. This finding suggests that a more parsimonious model relying primarily on meteorological, vegetation, and moisture variables could be evaluated in data-limited regions where

detailed infrastructure or socioeconomic datasets are unavailable or infrequently updated.

The relatively limited dependence on static geographic predictors also provides encouraging evidence regarding model transferability. Because much of the predictive signal originates from physically meaningful environmental processes rather than location-specific spatial memorization, the framework may transfer more readily than models dominated by static spatial predictors, provided that local wildfire observations and environmental data are available for calibration. Nevertheless, this hypothesis requires explicit validation. Future research should evaluate transferability using buffered spatial block cross-validation and completely independent geographic regions while incorporating dynamic anthropogenic variables, including seasonal agricultural burning practices, population dynamics, and land-use change, to better represent the human dimension of West African savannah fire regimes.

5. Conclusions

This study developed and evaluated a multi-source machine learning framework for predicting early-season wildfire risk in Ghana's Guinea Savannah by integrating meteorological, vegetation, soil, topographic, anthropogenic, and climate variables. Compared with a monthly Fire Weather Index (FWI) baseline, the machine learning models demonstrated substantially higher discriminative performance, indicating that integrating multiple environmental and human-related predictors provides a more informative representation of wildfire risk than weather variables alone. Statistical validation strengthened confidence in the proposed framework. Bootstrap confidence intervals demonstrated robust model performance, while pairwise DeLong tests showed that XGBoost significantly outperformed Random Forest, whereas differences between XGBoost and CatBoost and between CatBoost and Random Forest were not statistically significant. Probability calibration using isotonic regression substantially improved the reliability of predicted probabilities, highlighting the importance of calibration before operational implementation. Furthermore, bootstrap SHAP stability analysis (100 bootstrap resamples) produced a mean Spearman rank correlation of 0.983 ± 0.021 and a mean Kendall's Tau of 0.961 ± 0.050, indicating that feature-importance rankings were highly stable and reproducible.

The results demonstrate that wildfire occurrence in the Guinea Savannah is governed by the interaction of multiple climatic, environmental, and anthropogenic factors rather than by a single dominant predictor. Temperature, precipitation memory, vegetation condition, antecedent soil-moisture conditions, and human-related variables collectively contributed to model performance. These findings support the view that wildfire occurrence reflects the combined influence of fuel

availability, seasonal drying, and human ignition sources operating across space and time.

Spatial ablation analysis suggested that the model's predictive performance is not primarily driven by static spatial memorization. Removing the five most influential spatial variables reduced AUC by only 0.013, indicating that much of the predictive information originates from temporally varying environmental variables rather than location-specific characteristics. Although static variables such as roads, settlements, and topography contribute useful contextual information, the model derives most of its predictive capability from climate, vegetation, and moisture dynamics. This finding improves confidence in the framework's potential transferability while emphasizing the need for future evaluation using buffered or spatial block cross-validation.

The CMIP6 scenario analysis suggests that projected climatic warming could increase wildfire susceptibility across Ghana's Guinea Savannah, with larger increases under SSP5-8.5 than under SSP2-4.5. These results should be interpreted as scenario-based sensitivity analyses rather than operational forecasts, because future socioeconomic conditions were assumed to remain constant and tree-based machine learning models are inherently limited when extrapolating beyond the historical feature space. Future projection studies should incorporate dynamic socioeconomic pathways, land-use change, and additional climate models to better characterize uncertainty.

Although the proposed framework demonstrates strong predictive performance, it should currently be regarded as a research prototype for wildfire risk screening rather than a fully operational early-warning system. Operational implementation will require evaluation on the complete operational grid, calibration under true wildfire prevalence, threshold optimization for decision support, validation using independent fire observations, and testing under stricter spatial and temporal validation strategies. Additional comparisons with weather-only machine learning baselines using identical predictor availability would further clarify the benefits of the multi-source framework.

Several limitations should also be acknowledged. The monthly 1-km spatial resolution may not capture small fires or short-duration weather extremes, the Fire Weather Index baseline was evaluated at monthly rather than daily resolution, future projections relied on only two CMIP6 climate models, and anthropogenic variables were assumed to remain static under future climate scenarios. While bootstrap confidence intervals, DeLong significance tests, calibration

assessment, and SHAP stability analysis substantially reduced uncertainty regarding model performance, external validation in other fire-prone regions remains necessary before broader application.

Overall, this study demonstrates that integrating multi-source environmental and anthropogenic data within a machine learning framework provides a robust approach for early-season wildfire risk assessment in tropical savannah ecosystems. The framework offers improved predictive capability over conventional weather-only approaches while maintaining physically interpretable relationships between climate, vegetation, moisture, and human influences. With further prospective validation, operational calibration, and expanded climate-scenario analysis, the proposed framework provides a promising foundation for future wildfire risk assessment, climate adaptation planning, and evidence-based fire management across West African savannah landscapes

Funding

This research received no external funding.

Acknowledgments

Generative AI tools were used to support language editing. The author reviewed...

Declaration

The author declares no competing interests.

Data Availability

The raw source data used in this study are publicly available: MODIS fire and vegetation data from NASA FIRMS and LP DAAC (https://earthdata.nasa.gov), ERA5-Land climate data from ECMWF/Copernicus (https://cds.climate.copernicus.eu), CHIRPS precipitation from UCSB/CHG (https://www.chc.ucsb.edu/data/chirps), FLDAS soil moisture from NASA GMAO/GSFC, SRTM topography from USGS, and population data from WorldPop (https://www.worldpop.org). The processed 1-km gridded dataset including the 123 engineered features generated during this study are available from the corresponding author on reasonable request.

References

[1] D. M. J. S. Bowman et al., "Fire in the Earth System," Science, vol. 324, no. 5926, pp. 481–484, Apr.

2009, doi: 10.1126/science.1163886.
[2] W. Bond and J. Keeley, "Fire as a global 'herbivore': the ecology and evolution of flammable ecosystems," Trends in Ecology & Evolution, vol. 20, no. 7, pp. 387–394, Jul. 2005, doi: 10.1016/j.tree.2005.04.025.
[3] K. S. Dahan, R. A. Kasei, R. Husseini, M. Sarr, and M. Y. Said, "Analysis of the future potential impact of environmental and climate changes on wildfire spread in Ghana's ecological zones using a Random Forest (RF) machine learning approach," Remote Sensing Applications: Society and Environment, vol. 33, p. 101091, Jan. 2024, doi: 10.1016/j.rsase.2023.101091.
[4] E. E. Amoako and J. Gambiza, "Effects of fire on the population structure and abundance of Anogeissus leiocarpa and Vitellaria paradoxa in a West African savanna parkland," Acta Oecologica, vol. 112, p. 103745, Oct. 2021, doi: 10.1016/j.actao.2021.103745.
[5] L. K. M. Garnier and I. Dajoz, "The influence of fire on the demography of a dominant grass species of West African savannas, Hyparrhenia diplandra," Journal of Ecology, vol. 89, no. 2, pp. 200–208, Apr. 2001, doi: 10.1046/j.1365-2745.2001.00532. x.
[6] S. Archibald, D. P. Roy, B. W. Van Wilgen, and R. J. Scholes, "What limits fire? An examination of drivers of burnt area in Southern Africa," Global Change Biology, vol. 15, no. 3, pp. 613–630, Feb. 2009, doi: 10.1111/j.1365-2486.2008.01754. x.
[7] S. Catarino, J. M. N. Silva, A. K. Neves, D. Oom, and J. M. C. Pereira, "Fire weather severity in southern Africa is increasing faster and more extensively in the late than in the early dry season," International Journal of Wildland Fire, vol. 34, no. 3, Mar. 2025, doi: 10.1071/wf24002.
[8] C. E. Van Wagner, "Development and Structure of the Canadian Forest Fire Weather Index System," Canadian Forestry Service, Forestry Technical Report 35, Ottawa, Canada, 1987.
[9] J. Verbesselt, B. Somers, S. Lhermitte, I. Jonckheere, J. van Aardt, and P. Coppin, "Monitoring herbaceous fuel moisture content with SPOT VEGETATION time-series for fire risk prediction in savanna ecosystems," Remote Sensing of Environment, vol. 108, no. 4, pp. 357–368, Jun. 2007, doi: 10.1016/j.rse.2006.11.019.
[10] A. Issahaku, B. B. Campion, and R. Edziyie, "Rainfall and temperature changes and variability in the Upper East Region of Ghana," Earth and Space Science, vol. 3, no. 8, pp. 284–294, Aug. 2016, doi: 10.1002/2016ea000161.
[11] V. Resco de Dios et al., "A semi-mechanistic model for predicting the moisture content of fine litter," Agricultural and Forest Meteorology, vol. 203, pp. 64–73, Apr. 2015, doi: 10.1016/j.agrformet.2015.01.002.
[12] Y. Yu et al., "Machine learning–based observation-constrained projections reveal elevated global socioeconomic risks from wildfire," Nature Communications, vol. 13, no. 1, pp. 1–12, Mar. 2022, doi: 10.1038/s41467-022-28853-0.
[13] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/a:1010933404324.
[14] T. Chen and C. Guestrin, "XGBoost," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA: ACM, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[15] K. S. Dahan, R. A. Kasei, and R. Husseini, "Land use land cover (LULC) degradation intensity analysis of Guinea savannah and mosaic Forest savannah zones in Ghana," All Earth, vol. 35, no. 1, pp.

302–328, Sep. 2023, doi: 10.1080/27669645.2023.2255443.
[16] E. E. Amoako and J. Gambiza, "Effects of fire on the population structure and abundance of Anogeissus leiocarpa and Vitellaria paradoxa in a West African savanna parkland," Acta Oecologica, vol. 112, p. 103745, Oct. 2021, doi: 10.1016/j.actao.2021.103745.
[17] L. Giglio, W. Schroeder, and C. O. Justice, "The collection 6 MODIS active fire detection algorithm and fire products," Remote Sensing of Environment, vol. 178, pp. 31–41, Jun. 2016, doi: 10.1016/j.rse.2016.02.054.
[18] C. Funk et al., "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes," Scientific Data, vol. 2, no. 1, Dec. 2015, doi: 10.1038/sdata.2015.66.
[19] T. Dinku et al., "Validation of the CHIRPS satellite rainfall estimates over eastern Africa," Quarterly Journal of the Royal Meteorological Society, vol. 144, no. S1, pp. 292–312, Aug. 2018, doi: 10.1002/qj.3244.
[20] J. Muñoz-Sabater et al., "ERA5-Land: a state-of-the-art global reanalysis dataset for land applications," Earth System Science Data, vol. 13, no. 9, pp. 4349–4383, Sep. 2021, doi: 10.5194/essd-13-4349-2021.
[21] A. McNally et al., "A land data assimilation system for sub-Saharan Africa food and water security applications," Scientific Data, vol. 4, no. 1, Feb. 2017, doi: 10.1038/sdata.2017.12.
[22] A. Huete et al., "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," Remote Sensing of Environment, vol. 83, no. 1-2, pp. 195-213, Nov. 2002, doi: 10.1016/S0034-4257(02)00096-2.
[23] "MODIS/Aqua Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006," NASA Earthdata. [Online]. Available: https://www.earthdata.nasa.gov/data/catalog/lpcloud-myd15a2h-006
[24] T. G. Farr et al., "The Shuttle Radar Topography Mission," Reviews of Geophysics, vol. 45, no. 2, May 2007, doi: 10.1029/2005rg000183.
[25] J. R. Meijer, M. A. J. Huijbregts, K. C. G. J. Schotten, and A. M. Schipper, "Global patterns of current and future road infrastructure," Environmental Research Letters, vol. 13, no. 6, p. 064006, May 2018, doi: 10.1088/1748-9326/aabd42.
[26] M. Pesaresi et al., "The global human settlement layer from Landsat imagery," in 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, Jul. 2016, pp. 7276–7279, doi: 10.1109/igarss.2016.7730897.
[27] S. Panigrahi and V. K. Vidyarthi, "Assessing the suitability of McKee et al. (1993) drought severity classification across India," Natural Hazards, vol. 120, no. 14, pp. 13543–13572, Jul. 2024, doi: 10.1007/s11069-024-06762-3.
[28] S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, "A Multiscalar Drought Index Sensitive to Global Warming: The Standardised Precipitation Evapotranspiration Index," Journal of Climate, vol. 23, no. 7, pp. 1696–1718, Apr. 2010, doi: 10.1175/2009jcli2909.1.
[29] J. Castel-Clavera et al., "A Comparative Analysis of Fire-Weather Indices for Enhanced Fire Activity Prediction with Probabilistic Approaches," Elsevier BV, 2024, doi: 10.2139/ssrn.4856753.
[30] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna," in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, New York, NY, USA: ACM, Jul. 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.
[31] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in Advances in Neural Information Processing Systems 30, 2017, pp. 3146-3154.

[32] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in Advances in Neural Information Processing Systems 31, 2018, pp. 6638-6648.
[33] D. R. Roberts et al., "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," Ecography, vol. 40, no. 8, pp. 913-929, 2017, doi: 10.1111/ecog.02881.
[34] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems 30, 2017, pp. 4765-4774.
[35] B. C. O'Neill et al., "The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century," Global Environmental Change, vol. 42, pp. 169-180, Jan. 2017, doi: 10.1016/j.gloenvcha.2015.01.004.
[36] T. Mauritsen et al., "Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2," Journal of Advances in Modelling Earth Systems, vol. 11, no. 4, pp. 998–1038, Apr. 2019, doi: 10.1029/2018ms001400.
[37] S. Yukimoto et al., "The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and Basic Evaluation of the Physical Component," Journal of the Meteorological Society of Japan. Ser. II, vol. 97, no. 5, pp. 931–965, 2019, doi: 10.2151/jmsj.2019-051.
[38] K. Riahi et al., "The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview," Global Environmental Change, vol. 42, pp. 153-168, Jan. 2017, doi: 10.1016/j.gloenvcha.2016.05.009.
[39] D. Maraun, "Bias correcting climate change simulations - a critical review," Current Climate Change Reports, vol. 2, pp. 211-220, 2016, doi: 10.1007/s40641-016-0050-x.
[40] A. J. Cannon, S. R. Sobie, and T. Q. Murdock, "Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?" Journal of Climate, vol. 28, no. 17, pp. 6938–6959, Sep. 2015, doi: 10.1175/jcli-d-14-00754.1.
[41] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," Remote Sensing of Environment, vol. 202, pp. 18–27, Dec. 2017, doi: 10.1016/j.rse.2017.06.031.
[42] B. Ouattara et al., "Fire Impacts, vegetation Recovery, and environmental drivers in West African savannas (2014–2023): A High-Resolution remote sensing assessment," International Journal of Applied Earth Observation and Geoinformation, vol. 143, p. 104783, Sep. 2025, doi: 10.1016/j.jag.2025.104783.
[43] D. Entekhabi, E. Njoku, and P. O'Neill, "The Soil Moisture Active and Passive Mission (SMAP): Science and applications," in 2009 IEEE Radar Conference, IEEE, 2009, pp. 1–3, doi: 10.1109/radar.2009.4977030.
[44] N. Andela et al., "A human-driven decline in global burned area," Science, vol. 356, no. 6345, pp. 1356–1362, Jun. 2017, doi: 10.1126/science. aal4108.
[45] "Climate Change 2021: The Physical Science Basis," IPCC. [Online]. Available: https://www.ipcc.ch/report/ar6/wg1/.