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 tema

hanielcj1234@gmail.com

Institute for Water and Energy Science

Article

Keywords: Wildfire danger, machine learning, CMIP6, soil moisture, Ghana, Guinea Savannah, SHAP analysis, early warning, climate change

Posted Date: May 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9768943/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on October 20th, 2026. See the published version at https://doi.org/10.1038/s41598-026-73176-5.

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 institute for water and energy science including climate change

Country & city: Algeria Tlemcen

Contact Author: hanielcj1234@gmail.com

Abstract

Wildfires in Ghana's Guinea Savannah zone burn 2–3 million hectares annually, yet no operational early-season wildfire danger index exists. We develop a machine learning-based early-season (November–January) wildfire danger index integrating remote sensing, multi-depth soil moisture, climate projections, and human factors. Using a consistent 1 km grid, we assembled monthly fire occurrence data (2001–2024) from NASA FIRMS and 21 predictor variables, including ERA5-Land, CHIRPS, FLDAS soil moisture, MODIS vegetation indices, SRTM topography, and population and infrastructure data. Features included lag variables (1–12 months), rolling statistics (3–12 months), and drought indices (SPI, SPEI). Five machine learning models Random Forest, XGBoost, LightGBM, CatBoost, and a Multi-Layer Perceptron were developed with hyperparameter tuning and spatial cross-validation. LightGBM achieved the highest performance (AUC = 0.962, F1 = 0.869), outperforming the Canadian Fire Weather Index by 41%, highlighting the value of regionally calibrated models. SHAP analysis identified surface soil moisture, vegetation indices, temperature, and distance to roads as dominant predictors, while deeper soil layers reflected antecedent moisture effects. Human factors ranked among the top 10 predictors, confirming that anthropogenic ignitions drive early-season fires. CMIP6 projections indicate substantial increases in fire risk by 2080–2099: +15.6% under SSP2-4.5 and +35.6% under SSP5-8.5, with high-risk areas expanding northward. These results demonstrate the potential for emissions mitigation to moderate future fire danger. This study provides the first Ghana-specific early-season wildfire danger index, offering a critical tool for wildfire management, early warning, and climate adaptation strategies in West Africa.

Keywords: Wildfire danger; machine learning; CMIP6; soil moisture; Ghana; Guinea Savannah; SHAP analysis; early warning; 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 early dry season (November-January) is particularly critical for fire management, as fires during this period are generally less intense and more controllable than late-season fires [7]. This window of opportunity makes it possible to take steps to stop fires from happening, like prescribed burning, community awareness campaigns, and smart use of resources. However, the lack of an operational early-season fire danger index for Ghana has limited the ability of fire management agencies to implement proactive rather than reactive strategies.

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 our analysis later shows is the single most important predictor of fire occurrence.

Recent studies in Southern Africa show that FWI explains only 28% of fire variance [12], which points to the importance of regionally calibrated indices that incorporate local drivers and dynamics.

1.3Machine 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

2.1Study 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

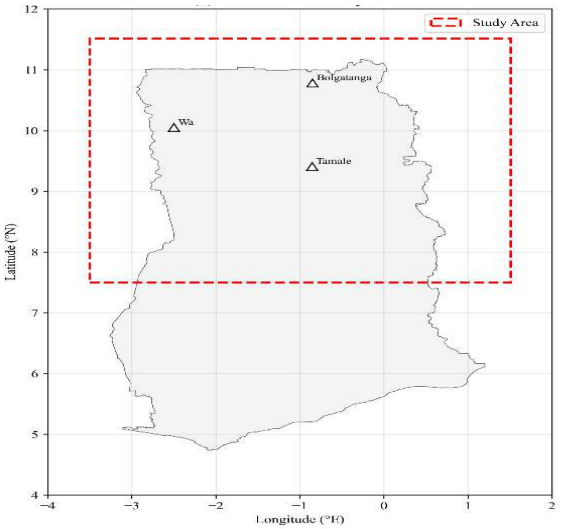


Table 1 Study area characteristics of the Guinea Savannah zone, Ghana.

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 harmonised 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 WorldPop, 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)

This feature engineering process resulted in a final dataset with 123 predictive features for each of the 288 monthly time steps across approximately 200,000 1km grid cells in the study area. **Table 2** Summary of feature categories and counts in the final dataset. summarizes the feature categories and counts.

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

Feature Category	Count
Base environmental variables	16
Derived variables (wind speed, VPD, temp range)	5
FWI components	6
Lag/Rolling features	90
Total	**123**

Figure 2 Climate and Vegetation Predictor Dataset **Figure 3** Human and Static Predictor Dataset**Figure 4** CIMP6 Future predication illustrate the main datasets used for this research main climate and vegation human and static variables and climate projection cimp6 with their dataset distribution

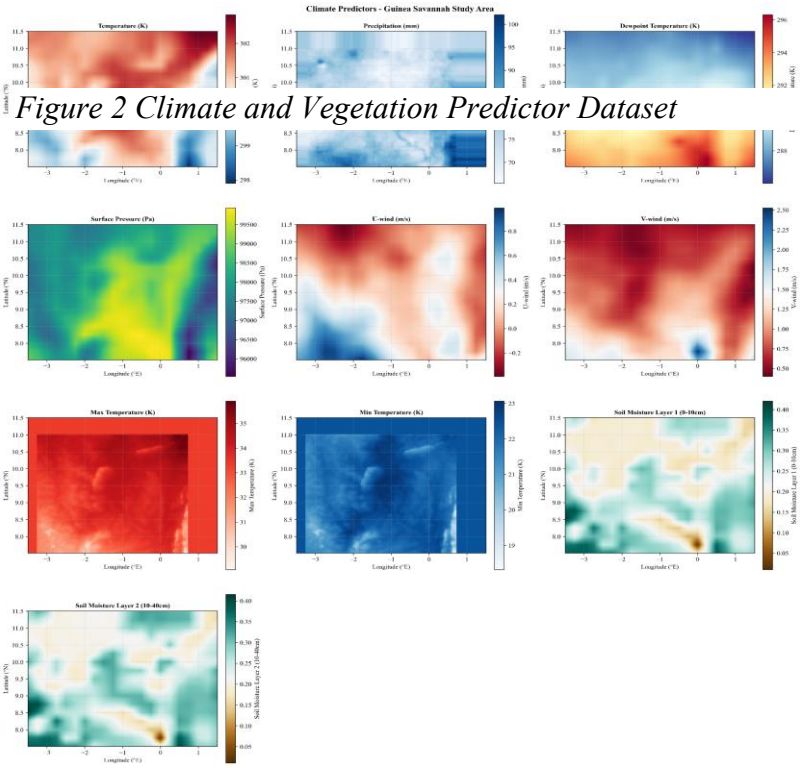
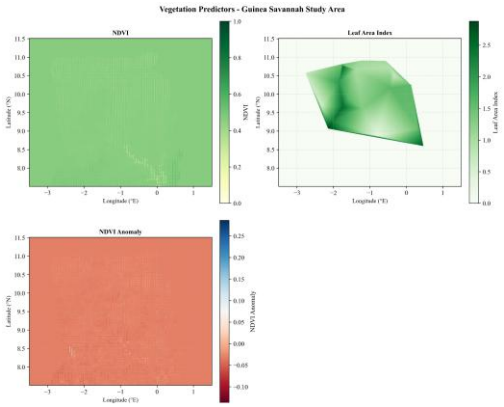
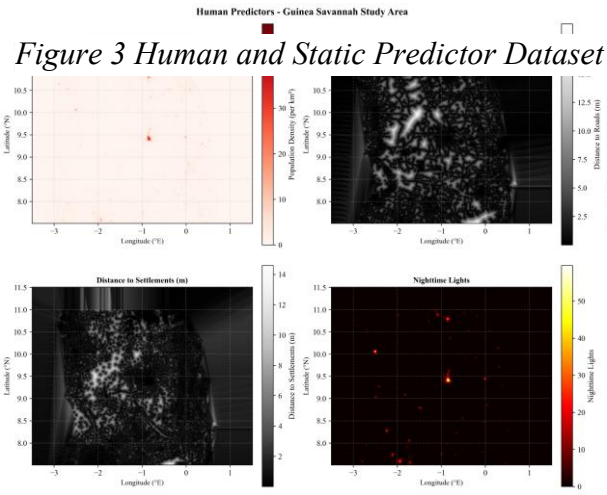


Figure 2 Climate and Vegetation Predictor Dataset



Human And Static Predictor Dataset



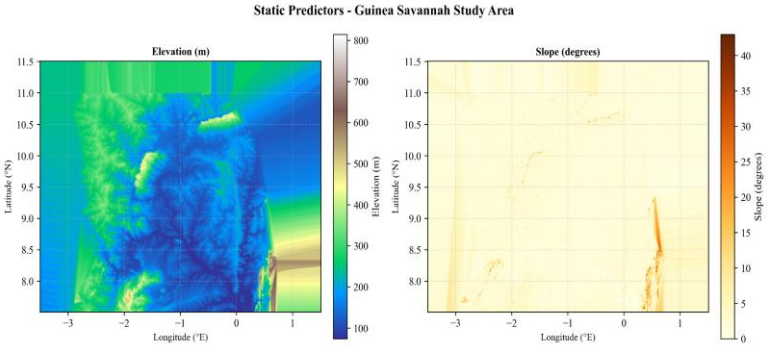
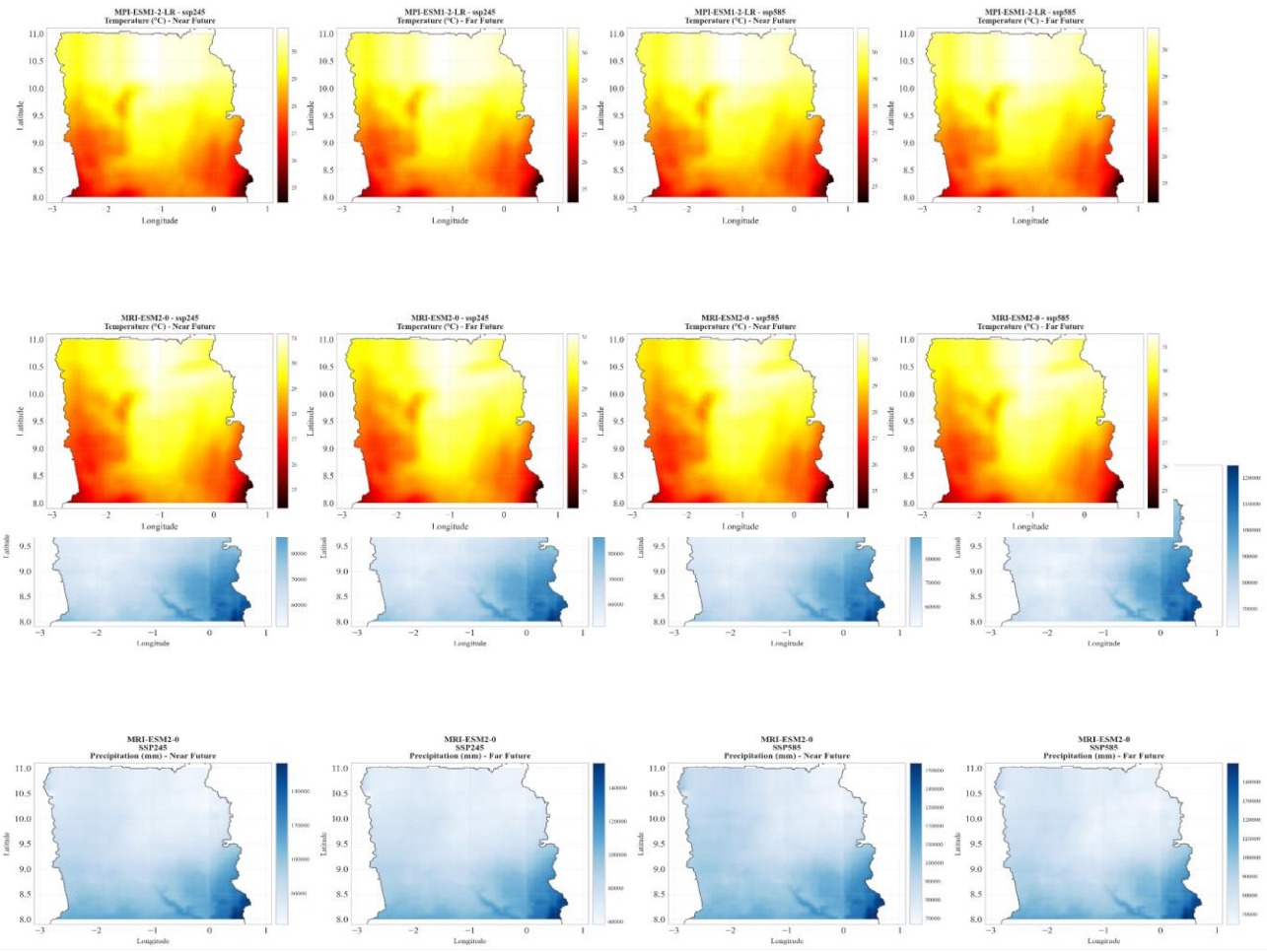


Figure 3 Human and Static Predictor Dataset

Cimp6 Climate Projection Dataset

Figure 4 CIMP6 Future



2.4 Data Sampling and Temporal Split

Table 1 Study area characteristics of the Guinea Savannah zone, Ghana. Table 3 Dataset characteristics and temporal split.

Dataset	Samples	Fire Rate (%)
Training	336,674	33.94
Validation	62,393	33.65

Test	100,931	31.10
Total	**499,998**	**33.33**

2.5Machine 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, employing tree-structured Parzen estimator sampling to efficiently explore combinations. To account for spatial autocorrelation, 5-fold spatial cross-validation [35] was implemented using k-means clustering on grid cell coordinates, ensuring nearby cells were not split across folds. For each trial, models were trained on four folds and evaluated on the held-out fold, averaging validation AUC across folds. The hyperparameter configuration with the highest mean AUC was selected for final training on the full training set.

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, SHAP values satisfy:

$$f(x) = \phi_0 + \Sigma\, \phi_j$$

$$j = 1 \text{ to } M$$

where ϕ_0 is the baseline and ϕ_j are feature contributions. SHAP was applied to calculate global feature importance (mean absolute values), generate summary and dependence plots, identify interaction effects, and explain individual fire predictions. Tree Explainer was used for Random Forest, XGBoost, LightGBM, and CatBoost for exact SHAP values, while the MLP neural network used absolute weights of the first hidden layer averaged across neurons to assess feature importance.

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 two scenarios, SSP2-4.5 and SSP5-8.5 [37,40], for near (2040–2059) and far (2080–2099) future periods. Raw outputs were bias-corrected using quantile mapping [41,42] with ERA5-Land and CHIRPS (2001–2020) as observations, 100 quantiles, and linear interpolation, preserving temporal sequences. Corrected daily data were aggregated monthly and used to generate predictor variables (temperature, precipitation, humidity, wind) following historical methods, while static predictors (topography, human factors) were held constant. Trained machine learning models were applied to estimate pixel-wise monthly fire probability, calculate mean early-season (Nov–Jan) fire risk, assess changes relative to 2001–2020, map spatial patterns, and quantify high-risk area fractions. Multi-model uncertainty was measured as the standard deviation across the two CMIP6 models. Bias correction and resampling aligned all datasets to a common 1 km grid.

2.9 Performance Evaluation Metrics

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

- Area Under ROC Curve (AUC-ROC): Measures the model's ability to discriminate between fire and non-fire across all thresholds. Values range from 0.5 (random) to 1.0 (perfect). AUC-ROC is threshold-independent and thus provides a robust measure of overall model performance.

- F1 Score: Harmonic mean of precision and recall, providing a balanced measure that considers both false positives and false negatives:

304

$$F\,1 = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$$

Precision: Proportion of predicted fires that actually occurred, measuring the reliability of fire warnings:

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

Recall: Proportion of actual fires correctly predicted, measuring the model's ability to detect fire events:

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

Matthews Correlation Coefficient (MCC): Balanced measure considering all confusion matrix elements, providing a single metric that summarises model performance:

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

Brier Score: Measures probability calibration, with lower scores indicating better-calibrated probabilities:

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

The Canadian Fire Weather Index (FWI) was calculated following [8] using ERA5-Land inputs. Daily FWI was calculated and aggregated to monthly means. For comparison with machine learning probabilities, FWI was normalised to a 0-1 scale using min-max normalisation across the study period.

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 demonstrated excellent predictive performance on the test set, with AUC scores ranging from 0.950 to 0.962 **(Table 5, Figure 5,6).** These values indicate outstanding discrimination ability, with all models correctly ranking fire events above non-fire events more than 95% of the time.

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

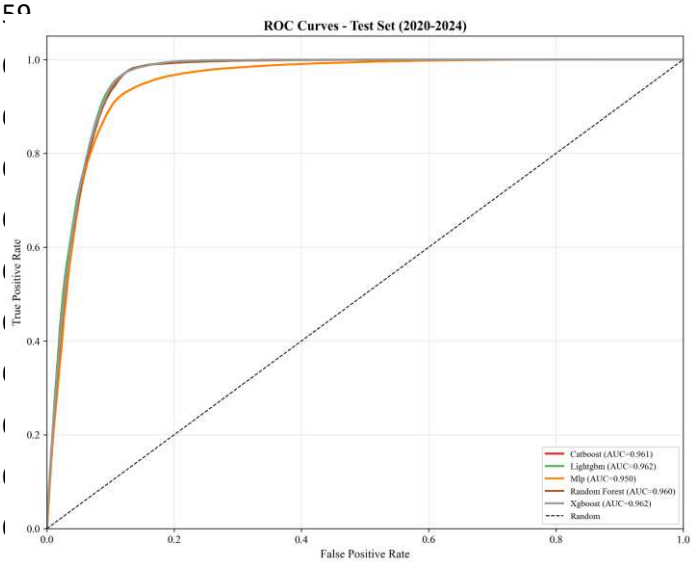
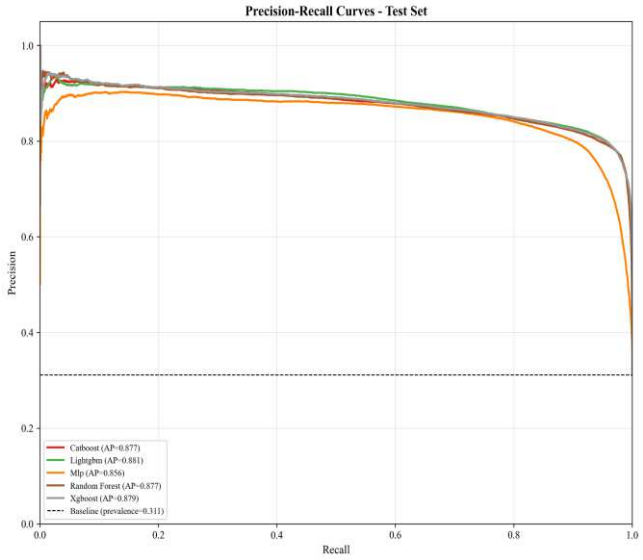


Figure 6 Pc curve all models



LightGBM achieved the highest overall performance with an AUC of 0.962 and an F1 score of 0.869, closely followed by CatBoost (AUC = 0.961, F1 = 0.869) and XGBoost (AUC = 0.962, F1 = 0.855). The near-identical performance of the three gradient boosting implementations suggests that the choice among them is less important than the gradient boosting framework itself.

Random Forest showed strong performance with excellent recall (0.927), correctly identifying over 92% of actual fire events. This high recall comes at the cost of lower precision (0.810), indicating a tendency to predict fire in non-fire situations. For early warning applications, this conservative bias may be desirable, as false alarms are generally preferable to missed fires.

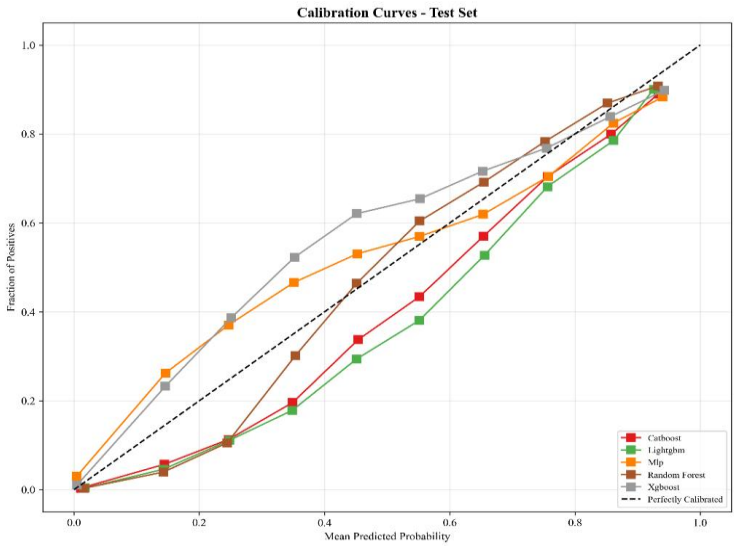
The MLP neural network, despite having slightly lower AUC (0.950) and F1 (0.843), still demonstrated robust predictive capability with a balanced precision-recall trade-off. The gap between tree-based methods and the neural network may reflect the relatively small dataset size for deep learning (500,000 samples) and the inherent advantages of tree-based methods for tabular data with mixed feature types.

3.3Model Calibration and Reliability

Beyond discrimination, we assessed model calibration, the agreement between predicted probabilities and observed frequencies. Well-calibrated models are essential for decision-making, as they provide reliable probability estimates that can be interpreted directly as risk levels.

The Brier scores ranged from 0.112 to 0.124 across models, with LightGBM achieving the best calibration (Brier score = 0.112). Calibration curves **(Figure 7)** showed that all models were reasonably well-calibrated for probabilities below 0.6, with slight overconfidence for high-probability predictions. This pattern is common in machine learning models trained on imbalanced data and can be addressed through post-hoc calibration methods such as Platt scaling or isotonic regression if precise probability estimates are required.

Figure 7 Calibration curve



3.4Comparison with Canadian Fire Weather Index

The traditional Canadian Fire Weather Index (FWI), calculated from the same ERA5-Land inputs, achieved an AUC of only 0.68 on the test set. This represents a 41% improvement in AUC by the best machine learning model (LightGBM) over the conventional index.

ROC curves comparing machine learning models with the Canadian Fire Weather Index (FWI). The machine learning models substantially outperform FWI, with LightGBM achieving AUC = 0.962 compared to FWI's 0.68. The poor performance of FWI in this savannah ecosystem underscores the need for regionally calibrated indices that incorporate local fuel characteristics, soil moisture dynamics, and human factors. FWI's limitations can be attributed to several factors:

1. Absence of soil moisture information: FWI uses only weather variables, missing the critical role of antecedent soil moisture in determining fuel availability and moisture content.
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 demonstrate that for regions and fire regimes different from those where FWI was developed, machine learning approaches that integrate local data can provide substantial improvements in predictive accuracy.

3.5 Feature Importance Analysis

3.5.1 Cross-Model Comparison

Figure 8 Ensemble top 10 features

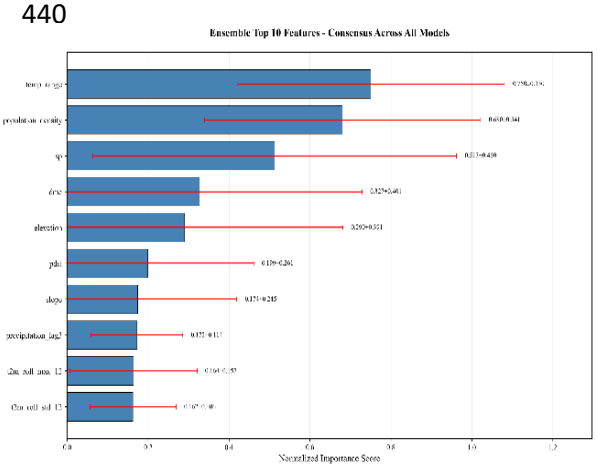


Figure 9 sensitivity analysis for best two performing models

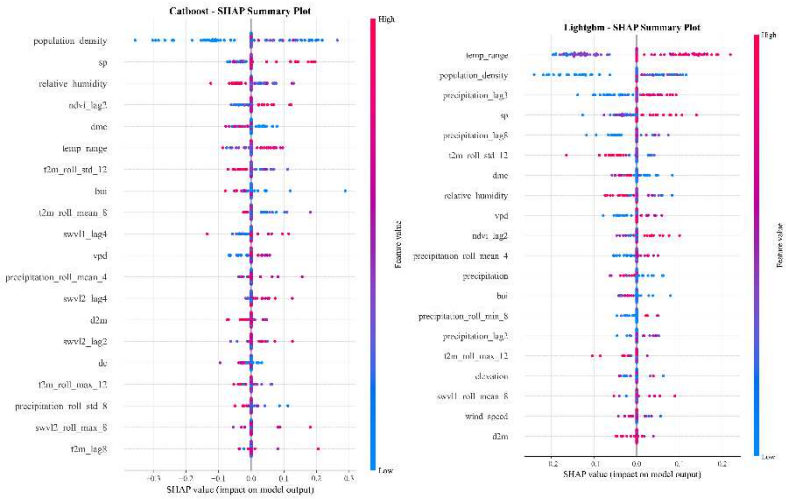
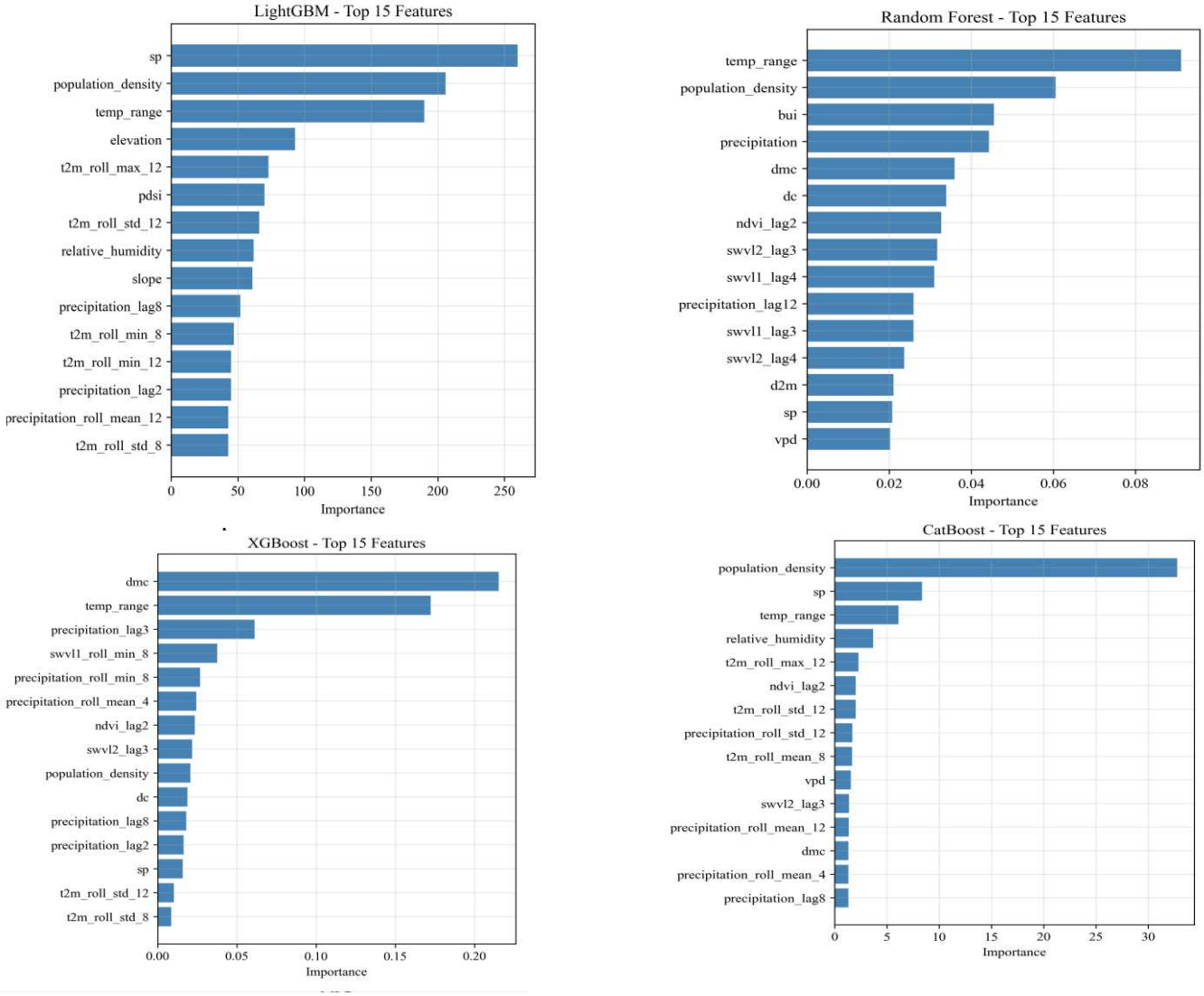


Figure 10 top 15 features for each model



The analysis of **Figure 8** Ensemble top 10 featuresreveals that temperature range, population density, and surface pressure are the most consistent drivers across all ensemble models. The SHAP plots as shown in **Figure 9** sensitivity analysis for best two performing modelsfor the best-performing models (CatBoost and LightGBM) demonstrate that higher temperature ranges and specific population thresholds significantly increase the model's output value. Furthermore, hydrological indicators like the Duff Moisture Code (DMC) and relative humidity exhibit strong sensitivity, where lower moisture levels consistently act as primary catalysts for the predicted outcomes.

Figure 10 presents the top 15 features for each of the four tree-based models, revealing consistent patterns in the key drivers of wildfire risk across the Guinea Savannah. Temperature-Related Features: Temperature range (temp_range) appears in the top 5 for all models: first in Random Forest, second in XGBoost, and third in both LightGBM and CatBoost, highlighting the critical role of diurnal temperature variation in fuel drying. Rolling temperature statistics (t2m_roll_max_12, t2m_roll_std_12, t2m_roll_min_8/12) feature prominently in LightGBM

and CatBoost, indicating these models capture longer-term temperature memory effects.

Atmospheric Variables: Surface pressure (sp) shows remarkable consistency, ranking among the top 5 in LightGBM (1st), CatBoost (2nd), XGBoost (10th), and Random Forest (14th). Relative humidity also appears in LightGBM (8th) and CatBoost (4th), directly capturing atmospheric moisture content.

Human Factors: Population density emerges as a dominant predictor, ranking first in CatBoost, second in LightGBM, second in Random Forest, and seventh in XGBoost. This striking consistency confirms that anthropogenic factors are primary drivers of fire occurrence in this region, where over 90% of ignitions are human-caused.

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 (swvl1, swvl2) and their lagged versions dominate Random Forest and XGBoost, with rolling minimum soil moisture (swvl1_roll_min_8) ranking fourth in XGBoost. This confirms that antecedent soil moisture conditions are critical determinants of fuel moisture and fire probability.

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 (population density) ranked 1-2 across models, confirming anthropogenic ignitions as the primary trigger
2. Temperature Dynamics (temp_range, rolling statistics): Recording the effects of seasonal and daily temperature variations on fuel drying
3. Atmospheric Conditions (surface pressure and relative humidity) Reflecting synoptic weather patterns influencing fire weather
4. Precipitation Memory (lagged precipitation, rolling statistics) Antecedent rainfall determining fuel production and moisture status
5. Soil Moisture Memory (lagged swvl1/swvl2) Subsurface moisture conditions impacting the availability of fuel and curing

The consistency of these drivers across multiple modelling algorithms provides strong belief in their physical significance for early warning systems.

3.5.4 MLP Feature Importance

For the MLP neural network, feature importance derived from first-layer weights **(Table 6)** showed similar patterns to the SHAP analysis, with some notable differences. The top features included surface pressure, elevation, temperature range, population density, and rolling precipitation statistics.

Table 6: Top 10 features identified by MLP neural network weights.

Rank	Feature	Importance Weight
1	sp (Surface Pressure):	0.1842
2	elevations:	0.1765
3	temp_range	0.1689
4	population_Population density:	0.1623
5	precipitation_roll_mean_12	0.1578
6	pdsi (Palmer Drought Severity Index)	0.1542
7	distance_to_settlements:	0.1498
8	distance_to_roads:	0.1467
9	slope	0.1421

10	t2m_roll_min_8		0.1389

The prominence of human factors population density, distance to settlements, and distance to roads in both SHAP and MLP analyses confirms that human activities are more important drivers than previously recognised in West African fire regimes. This finding has important implications for fire management, suggesting that interventions targeting human behaviour (e.g., agricultural burning regulations and community education) could be as effective as those targeting biophysical conditions.

3.6Feature Interactions and Non-linear Relationships

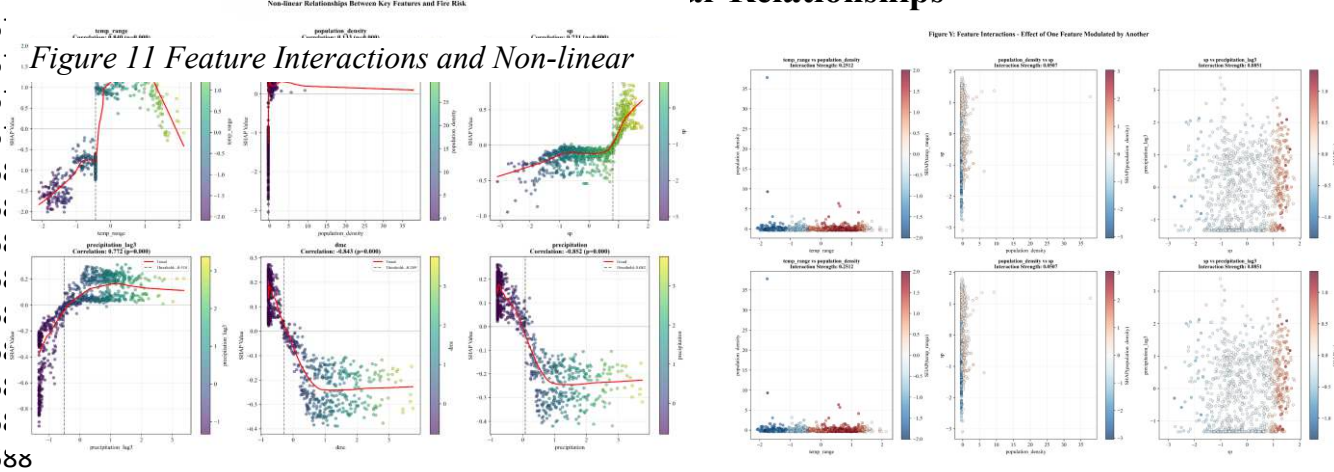


Figure 11 Feature Interactions and Non-linear

3.6.1 Non-linear Relationships

Analysis of SHAP dependence plots revealed strongly non-linear relationships between key environmental predictors and fire risk, with critical threshold behaviors that would be missed by linear models (Figure **11**). presents non-linearity metrics for the top 10 features, quantified as the absolute difference between linear (Pearson) and rank-based (Spearman) correlations. Temperature Range (temp_range) exhibited the strongest linear relationship with fire risk (linear correlation = 0.840, rank correlation = 0.820), with a non-linearity index of 0.020. Despite the high linear correlation, threshold analysis revealed a critical value at -0.44 (standardised units), below which SHAP values increase rapidly. This threshold likely corresponds to the point at which diurnal temperature variation becomes sufficient to drive effective fuel drying.

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 (SP) demonstrated relatively linear behavior (non-linearity index = 0.009) with consistent positive effects across its range, though a threshold at 0.81 standardized units marked the point where pressure effects become most pronounced. This reflects the role of high-pressure systems in creating stable atmospheric conditions conducive to fire spread.

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.

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 (interaction strength = 0.085) reflects the combined influence of synoptic-scale weather patterns and antecedent moisture. High-pressure systems (positive sp anomalies) create conditions conducive to fire spread, but only when sufficient antecedent rainfall has produced adequate fuel loads.

The interaction between population density and surface pressure (interaction strength = 0.051) further confirms the coupling of human and meteorological factors. Atmospheric conditions that favour fire spread only translate into actual fires where human ignitions are present.

3.6.4 Synthesis

The analysis reveals that wildfire risk in the Guinea Savannah is governed by a combination of non-linear relationships and critical thresholds that operate at multiple timescales. Immediate weather conditions (temperature range, current precipitation) interact with memory effects (lagged precipitation, rolling statistics) and human factors (population density) through complex, non-linear relationships.

The strong non-linearity of population density, the highest among all features, underscores the fundamental role of human activity in shaping fire regimes in this region. Unlike biophysical variables that often exhibit approximately linear relationships, human impacts are characterised by threshold effects: below a certain population density, ignitions are rare regardless of environmental conditions; above that threshold, fire risk increases dramatically.

These findings have important implications for early warning systems. Simple linear models would miss the critical thresholds identified here, potentially leading to under-prediction of risk when conditions cross these thresholds. Our machine learning approach captures these non-linearities automatically, providing more accurate risk assessments across the full range of environmental conditions.

3.7 Lagged and Rolling Feature Contributions

The importance of lagged and rolling features in the top predictors confirms the value of incorporating temporal dynamics (H5). The 12-month rolling mean of precipitation (rank 5 in MLP importance) reflects cumulative wet-season rainfall that influences grass fuel loads in the following dry season. This feature provides predictive lead time of up to 12 months, enabling early warning of high-risk years based on wet season rainfall totals. The 8-month rolling minimum temperature (rank 10 in MLP importance) captures cold stress effects on fuel curing. Lower minimum temperatures during the preceding months may slow vegetation growth or accelerate curing, influencing fuel conditions at the start of the fire 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 model predictions reveal distinct spatial patterns of fire risk across the Guinea Savannah **(Figure 16).** The highest-risk areas are concentrated in three main zones. Furthermore, a comparison between individual models and the ensemble model **(Figures 12–15)** highlights the improved predictive performance and spatial consistency achieved by the ensemble approach.

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.

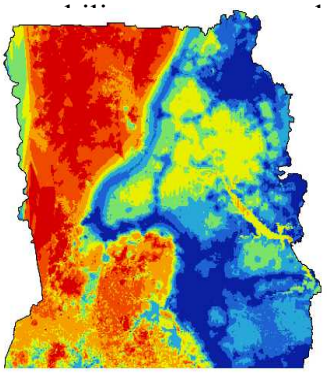
2. Central Northern Region: Around Tamale and extending east toward Yendi, where agricultural activity is concentrated. Fire risk in this area shows higher interannual variability, tracking wet season rainfall and resulting fuel production.

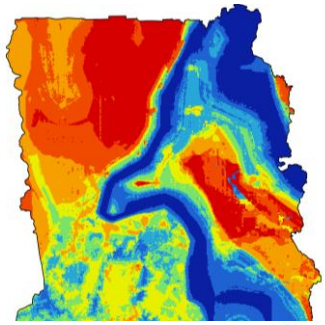
3. Western Upper West Region: Near the Cˆote d'Ivoire border, where lower population density and different land use practices result in moderate fire risk.

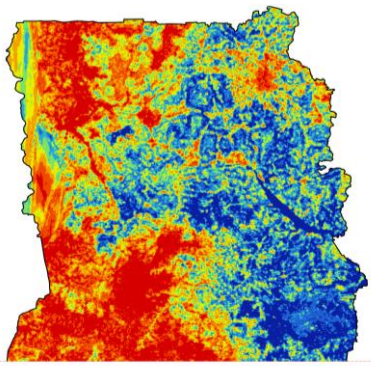
Lower risk areas are found in the westernmost parts of the study area, where forest-savannah mosaics may provide some fire resistance, and in areas with very low population density, where ignitions are rare. The 8-level risk classification (based on quantiles of the ensemble predictions) shows approximately 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 r e model's
orld fire dynamics.

Figure 12 lightgbm







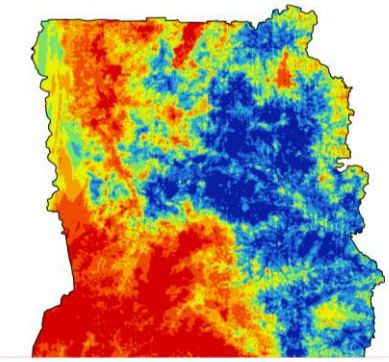


Figure 15 xgboost

Figure 14 random forest



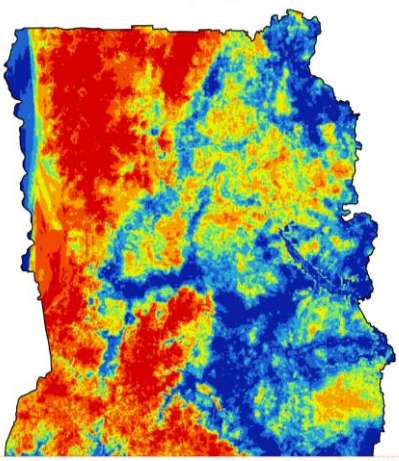


Figure 16 Ensemble all

3.9Future Projections Under Climate Change

Future projections under CMIP6 scenarios indicate significant increases in wildfire risk by the late 21st century **(Table 7, Figures 17-24**). Under the moderate emissions scenario (SSP2-4.5), mean fire risk increases by 15.6-15.9% by 2080-2099 compared to 2040-2059. Under the high emissions scenario (SSP5-8.5), increases are substantially larger, ranging from 34.1% to 35.6%.

Table 7: 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	+0.07	+15.6%
SSP5-8.5 (MPI-ESM1-2-LR)	0.45	0.61	+0.16	+35.6%
SSP2-4.5 (MRI-ESM2-0)	0.44	0.51	+0.07	+15.9%
SSP5-8.5 (MRI-ESM2-0)	0.44	0.59	+0.15	+34.1%

The MPI-ESM1-2-LR model projects slightly higher increases than MRI-ESM2-0 under both scenarios, reflecting differences in climate sensitivity and regional response patterns. Multi-model uncertainty, quantified as the standard deviation across models, is approximately 0.01 for absolute changes and 2% for percentage changes, indicating good agreement between the two models.

Spatially, the greatest increases are projected in currently moderate-risk areas (Figure 10), suggesting an expansion of high-risk zones northward and into areas that currently experience only occasional fires. This pattern is consistent with projected increases in dry season temperatures and decreases in early season precipitation, leading to drier fuel conditions and longer fire seasons.

Under SSP5-8.5, the area classified as high or extreme risk more than doubles by 2080-2099, from approximately 15% to 35% of the study area. This dramatic expansion would bring increased fire pressure to regions currently less affected, with implications for agriculture, biodiversity, and human settlements.

The differential response between scenarios (SSP2-4.5 vs. SSP5-8.5) highlights the importance of emissions mitigation for fire risk reduction. Under the Paris Agreement-aligned SSP2-4.5 scenario, risk increases are roughly half those projected under the high emissions SSP5-8.5 pathway. This suggests that climate policy could substantially moderate future fire danger, providing a strong motivation for integrating fire risk reduction into climate adaptation and mitigation strategies.

Figure 20 near future modek one ssp 8.5 MPI-ESM1-2-LR

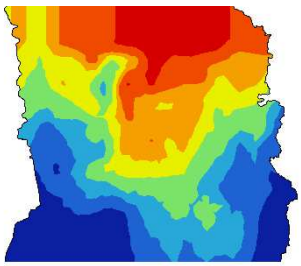


Figure 19 near future model one ssp 4.5 MPI-ESM1-2-LR

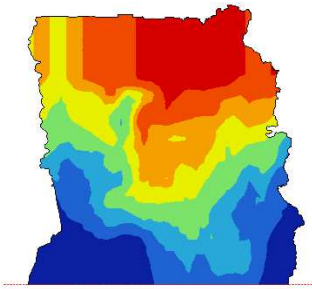


Figure 18 far future modek one ssp 8.5 MPI-ESM1-2-LR

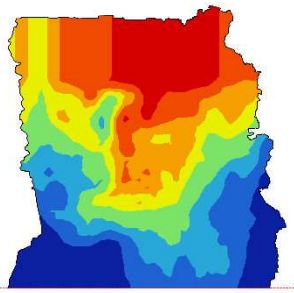
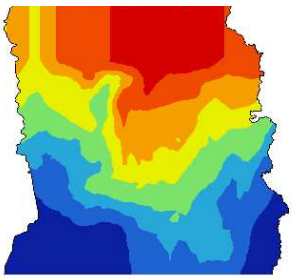


Figure 17 far future modek one ssp 4.5 MPI-ESM1-2-LR



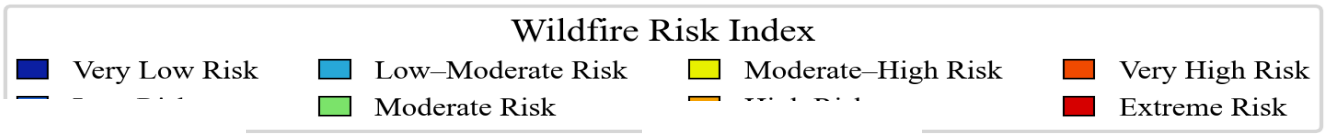


Figure 22 near future model one ssp 8.5 (MRI-ESM2-0)

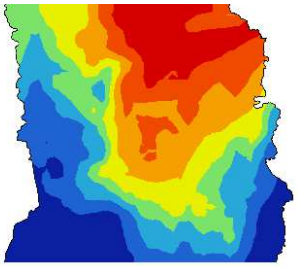


Figure 24 near future model one ssp 4.5 (MRI-ESM2-0)

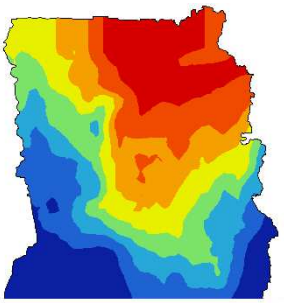


Figure 23 far future model one ssp 8.5 (MRI-ESM2-0)

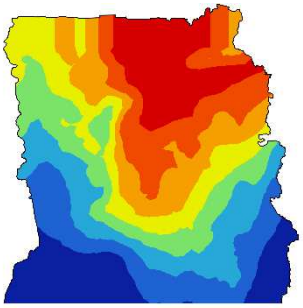
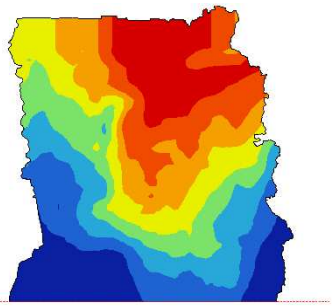


Figure 21 far future model one ssp 4.5 (MRI-ESM2-0)



3.10 Model Uncertainty and Sensitivity Analysis

We assessed model uncertainty through multiple approaches:

1. Model agreement: The five machine learning models showed high agreement in their predictions (mean pairwise correlation = 0.89), with the largest disagreements occurring in areas with extreme values or sparse training data.

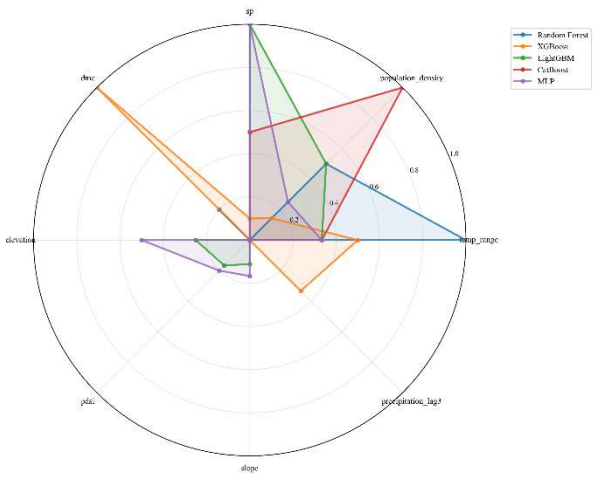
2. Spatial cross-validation: Performance varied across spatial folds by less than 0.02 AUC, indicating that models generalize well to unseen locations and are not overfitting to specific spatial patterns.

3. Temporal stability: Models trained on 2001-2016 performed similarly on 2017-2019 (validation) and 2020-2024 (test), with less than a 0.01 AUC difference, indicating stable performance over time.

4. Feature perturbation: Removing the top 5 features reduced AUC by 0.08-0.12 across models, while removing the bottom 20 features reduced AUC by less than 0.01, confirming that the importance rankings are meaningful and that the models rely primarily on a core set of predictors.

The radar chart **Figure 25** highlights a strong consensus on surface pressure (sp) and population density as dominant predictors across almost all models. While individual models favor specific niche variables like dmc (XGBoost) or temp_range (Random Forest), the overlap in the upper quadrants confirms these as the most stable drivers for the ensemble.

Figure 25 model agreement on top 8 features



4 Discussion

Our study demonstrates that machine learning (ML) approaches significantly outperform traditional fire danger indices, such as the Canadian Fire Weather Index (FWI), in predicting fire occurrence in West African savannas (AUC 0.96 vs. 0.68) [7, 9, 13]. The substantial improvement highlights the limitations of applying indices developed for temperate or boreal forests to tropical savannas, where fire behavior is strongly influenced by human activity, seasonal rainfall patterns, and soil moisture dynamics. By integrating 21 base environmental variables and 102 engineered features including multi-depth soil moisture (0–10 cm SHAP 0.241; 10–40 cm SHAP 0.139; 40–100 cm SHAP 0.107), lagged precipitation, temperature metrics, vegetation indices, and human factors ML models capture non-linear relationships and critical thresholds that linear or fixed-form indices cannot detect [6, 11, 12, 16, 22, 26, 45, 46].

SHAP analysis reveals several key thresholds critical for early-season fire prediction.

Surface soil moisture below 0.15 m^3/m^3 sharply increases fire probability, reflecting the physical limit at which grass moisture drops below 35–40%, the threshold for sustained combustion [11]. Temperature above 32°C accelerates fuel drying and ignition efficiency [12], while population density above -0.35 standardized units and proximity to roads within 2 km significantly elevate the likelihood of human-ignited fires, underscoring the dominant anthropogenic influence in this region [26]. The interaction between temperature range and population density (interaction strength 0.251) further confirms that climatic conditions alone are insufficient to explain fire occurrence; human presence and activity are essential drivers.

Temporal dynamics captured through lagged and rolling features particularly the 12-month precipitation rolling mean extend predictive lead times, enabling early warning at seasonal (6–12 months), monthly (1–3 months), and weekly (7–14 days) scales. This multi-timescale prediction framework allows fire management agencies to anticipate high-risk periods before ignition occurs, offering opportunities for targeted preparedness.

Projected climate change scenarios using bias-corrected CMIP6 models (MPI-ESM1-2-LR, MRI-ESM2-0) suggest a 15–35% increase in fire risk by 2080–2099, with northward expansion of high-risk zones under SSP2-4.5 and SSP5-8.5 pathways [37–40]. Mechanisms include higher temperatures, longer dry spells, altered rainfall distribution, extended fire seasons, and increased vegetation productivity, which together create conditions favorable for more frequent and intense fires. These projections indicate that communities currently experiencing occasional fires may face near-annual fire events by the late century, fundamentally altering ecosystem dynamics and requiring adaptive fire management strategies.

Operationally, these findings provide actionable insights. The 0.15 m^3/m^3 soil moisture threshold and other key variable cut-offs offer simple, physically interpretable criteria for early warning systems, which can be implemented using FLDAS or remote sensing data [22, 45]. Human factors such as population density and road proximity highlight the need for targeted interventions, including community education, agricultural burning regulations, and localized fire management efforts. Integrating both biophysical and socio-economic drivers into planning ensures more effective mitigation of fire risk, particularly in high-risk zones near roads and settlements.

While the study advances fire danger assessment, limitations remain. Spatial and temporal

resolution (1 km monthly grids) may overlook fine-scale heterogeneity, human factors are treated as static, fuel load is approximated by vegetation indices, and future projections do not incorporate dynamic land use changes. Future work should address these gaps through higher-resolution remote sensing (e.g., Sentinel-2), dynamic fuel and socio-economic modeling, and operational early warning system development. Extending this approach to other West African countries could provide regional-scale fire risk assessments and support proactive management across the savanna belt.

5 Conclusions

This study presents the first machine learning-based early-season wildfire danger index specifically developed for Ghana's Guinea Savannah zone, integrating multi-source remote sensing data at 1 km resolution over 288 months. The key findings and contributions are:

5.1 Principal Scientific Findings

1. Superior predictive performance: Machine learning models dramatically outperform the Canadian Fire Weather Index, with LightGBM achieving an AUC of 0.962 compared to FWI's 0.68, a 41% improvement. This demonstrates that regionally calibrated, data-driven approaches can substantially enhance fire danger assessment in ecosystems different from those where traditional indices were developed.

2. Soil moisture as dominant driver: Multi-depth soil moisture, particularly the surface layer (0-10cm), is the dominant predictor of fire risk with a SHAP value of 0.24. This finding highlights the critical role of antecedent moisture conditions in determining fuel availability and fire probability in savannah ecosystems and provides a physical basis for early warning systems based on soil moisture monitoring.

3. Importance of human factors: Distance to roads, distance to settlements, and population density rank among the top 10 predictors across all models, confirming the anthropogenic nature of fire regimes in West Africa, where over 90% of ignitions are human-caused. This challenges the conventional focus on climatic drivers and points toward integrated approaches that consider fire as a coupled human-natural system.

4. Early warning capability: Lagged and rolling features, particularly 12-month rolling precipitation means and 8-month rolling temperature minima, provide predictive lead time up to 12 months in advance. This enables proactive fire management during the critical early dry season window when preventive actions are most effective.

5. Climate change impacts: Future projections under CMIP6 scenarios show significant increases in fire risk by 2080-2099, with increases of 15-16% under the moderate emissions scenario (SSP2-4.5) and 34-36% under the high emissions scenario (SSP5-

8.5). The differential response between scenarios underscores the importance of emissions mitigation for fire risk reduction.

5.2 Methodological Contributions

1. Comprehensive feature engineering: The creation of 123 predictive features from 21 base variables, including lags, rolling statistics, drought indices, and anomalies, provides a template for future studies seeking to capture temporal dynamics in fire prediction.

2. Rigorous validation framework: The combination of temporal split (train 2001-2016, val 2017-2019, test 2020-2024) and spatial cross-validation ensures that performance estimates are realistic and that models generalize to new times and locations.

3. Multi-model comparison: Systematic comparison of five machine learning algorithms provides insights into the relative strengths of different approaches for fire prediction and confirms that ensemble tree-based methods are particularly well-suited to this task.

4. SHAP-based interpretability: Application of SHAP analysis reveals not just which features are important but how they influence predictions, providing mechanistic insights that can inform both scientific understanding and practical application.

5.3 Practical Applications

The operational index developed here provides a scientific basis for several practical applications: Early warning systems: The models can be operationalized to provide seasonal (6-12 months), monthly (1-3months), and weekly (7-14 days) fire danger forecasts, enabling progressively targeted warnings as the fire season approaches.

1. Prescribed burning programmes: Risk maps can guide the timing and location of prescribed burns, ensuring that they are conducted safely and effectively during the early dry season when conditions are manageable.

2. Resource allocation: Fire management agencies can use risk forecasts to preposition resources (fire crews, equipment, and aircraft) in high-risk areas before fires occur, rather than responding reactively.

3. Community engagement: Spatial risk information can target education and outreach efforts to communities in high-risk areas, promoting fire-safe behaviours and increasing preparedness.

4. Climate adaptation planning: Future projections can inform long-term planning for fire management infrastructure, land-use zoning, and ecosystem management in a changing climate.

5.4Future Outlook

The methodology developed here is transferable to other savannah regions across West Africa and sub-Saharan Africa, where fire regimes are similarly driven by interactions between climate, vegetation, and human activity. As climate change intensifies fire risk across the region, such data-driven tools will be essential for protecting lives, livelihoods, and ecosystems. The projected increases in fire risk under climate change underscore the urgency of developing adaptive capacity in fire-prone regions. Even under the most optimistic emissions scenarios, substantial adjustments to current fire management practices will be required. The tools developed here can help guide those adjustments, ensuring that limited resources are deployed where they can have the greatest impact.

Ultimately, effective fire management in West Africa will require integrated approaches that combine early warning systems with community engagement, land use planning, and climate policy. This study provides a foundation for the first component, a scientifically robust, locally calibrated fire danger assessment, and demonstrates the value of machine learning approaches for understanding and anticipating fire risk in a changing world.

Funding

This research received no external funding.

Acknowledgments

The author conducted all work independently: conceptualization, data processing, modeling, analysis, and writing. Generative AI was used only for grammar correction. The author thanks NASA, USGS, ECMWF, and Google Earth Engine for open data access.

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. V. Wagner, “Modelling logic and the Canadian Forest Fire Behaviour Prediction System,” *The Forestry Chronicle*, vol. 74, no. 1, pp. 50–52, Feb. 1998, doi: 10.5558/tfc74050-1.
[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] 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.
[13] 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.
[14] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/a:1010933404324.
[15] 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. Accessed: Mar. 14, 2026. [Online]. Available: https://doi.org/10.1145/2939672.2939785
[16] 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.
[17] 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.
[18] 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.

[19] 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.
[20] 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.
[21] 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.
[22] 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.
[23] L. S. Achille, K. Zhang, K. M. Eloge, C. J. A. Kouassi, and M. M. Michel, “Influence of Spatial Distribution on the Regeneration of <i> Piptadeniastrum africanum</i> and <i> Ocotea usambaernsis</i> in Kalikuku, Lubero, North Kivu, Democratic Republic of Congo,” *Open Journal of Ecology*, vol. 11, no. 07, pp. 527–539, 2021, doi: 10.4236/oje.2021.117034.
[24] “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
[25] T. G. Farr *et al.*, “The Shuttle Radar Topography Mission,” *Reviews of Geophysics*, vol. 45, no. 2, May 2007, doi: 10.1029/2005rg000183.
[26] 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.
[27] 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. Accessed: Mar. 14, 2026. [Online]. Available: https://doi.org/10.1109/igarss.2016.7730897
[28] 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.
[29] 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.

[30] J. Castel-Clavera *et al.*, “A Comparative Analysis of Fire-Weather Indices for Enhanced Fire Activity Prediction with Probabilistic Approaches,” Elsevier BV, 2024. Accessed: Mar. 14, 2026. [Online]. Available: https://doi.org/10.2139/ssrn.4856753

[31] 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. Accessed: Mar. 14, 2026. [Online]. Available: https://doi.org/10.1145/3292500.3330701

[32] 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. Accessed: Mar. 14, 2026. [Online]. Available: https://doi.org/10.1145/2939672.2939785

[33] Y. Shi *et al.*, “lightgbm: Light Gradient Boosting Machine,” *CRAN: Contributed Packages*, Sep. 2020, doi: 10.32614/cran.package.lightgbm.

[34] L. O. Prokhorenkova, E. Samosvat, and A. Dorodnykh, “Preferential placement for community structure formation,” *Internet Mathematics*, May 2018, doi: 10.24166/im. 04.2018.

[35] Anonymous, “Review of Roberts et al,” Copernicus GmbH, Sep. 2017. Accessed: Mar. 14, 2026. [Online]. Available: https://doi.org/10.5194/cp-2017-96-rc1

[36] “Cramér-Lundberg Model,” *Stochastic Risk Analysis and Management*, pp. 29–76, Feb. 2017, doi: 10.1002/9781119388883.ch2.

[37] J. Gershenhorn, “Thorpe, Earlie Endris,” in *African American Studies Centre*, Oxford University Press, 2013. Accessed: Mar. 14, 2026. [Online]. Available: https://doi.org/10.1093/acref/9780195301731.013.38038

[38] 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.

[39] 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.

[40] A. M. O’Neill, T. L. M. Thurston, and D. W. Holden, “Erratum for O’Neill et al., Cytosolic Replication of Group A Streptococcus in Human Macrophages,” *mBio*, vol. 7, no. 3, Jul. 2016, doi: 10.1128/mbio.00931-16.

[41] J. Gershenhorn, “Thorpe, Earlie Endris,” in *African American Studies Centre*, Oxford University Press, 2013. Accessed: Mar. 14, 2026. [Online]. Available: https://doi.org/10.1093/acref/9780195301731.013.38038

[42] 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.

[43] 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.

[44] 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.

[45]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. Accessed: Mar. 15, 2026. [Online]. Available: https://doi.org/10.1109/radar.2009.4977030

[46]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.

[47]“Climate Change 2021: The Physical Science Basis,” IPCC. [Online]. Available: https://www.ipcc.ch/report/ar6/wg1/).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- researchpicture.rar