World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(Research Article)

Integrating Multi-Resolution Remote Sensing Data for Daily Forest Fire Risk Forecasting in a Nigerian Savanna Ecosystem

Oleka Chidubem Meshack, Paschal Chinedu Ohiri, Eleazar Benson Mfonobong *, Chrysogonus Chukwumere Ogomaka and Chika Juliana Anyalewechi

Department of Electrical and Electronic Engineering, Federal University of Technology, Owerri, Imo State, Nigeria.

World Journal of Advanced Engineering Technology and Sciences, 2026, 18(02), 263-278

Publication history: Received on 13 January 2026; revised on 21 February 2026; accepted on 23 February 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.18.2.0106

Abstract

Forest fires represent a persistent environmental and socio-economic threat across West African savanna ecosystems, where limited ground-based monitoring and persistent cloud cover constrain effective early-warning systems. Remote sensing offers a powerful alternative for wildfire monitoring; however, the trade-off between high spatial resolution and high temporal frequency remains a key challenge for operational fire risk forecasting. This study developed a pragmatic spatio-temporal data fusion framework for daily forest fire risk forecasting using multi-resolution remote sensing and meteorological data, with Kainji Lake National Park, Nigeria, as a case study. High-resolution vegetation indices derived from Sentinel-2 imagery were integrated with daily MODIS surface reflectance products and ERA5-Land meteorological reanalysis data within the Google Earth Engine platform to generate a continuous daily dataset at a harmonized spatial resolution. VIIRS active fire detections was deployed, enabling the formulation of forest fire forecasting as a temporal classification problem. Random Forest and Extreme Gradient Boosting (XGBoost) models were trained. Both models demonstrated strong predictive performance on an independent test dataset, achieving high discrimination between fire and no-fire days. Random Forest slightly outperformed XGBoost, attaining an area under the receiver operating characteristic curve of 0.997 and an F1-score of 0.957, while both models achieved perfect recall for fire events. The results highlighted that daily fire risk in Kainji Lake National Park was primarily governed by seasonal and atmospheric conditions rather than vegetation greenness alone. The proposed framework provides a scalable foundation for early-warning systems and fire management applications in Nigeria and similar regions across West Africa.

Keywords: Forest Fire; MODIS; ERA5-Land; Kainji; Early-Warning; Xgboost; VIIRS

1. Introduction

Forest fires pose a significant and growing threat to ecosystems, biodiversity, and human livelihoods across the globe, including in Nigeria [1]. They drive loss of tree cover, degrade biodiversity, and contribute to air pollution and carbon emissions, with remote-sensing records showing notable amounts of fire-driven tree-cover loss across the country in recent decades [2]. Covering approximately 20.4 million hectares of natural forest in 2020—equivalent to 22% of its land area—Nigeria has experienced accelerating tree cover loss, with 207,000 hectares deforested in the same year, releasing an estimated 87.4 million tons of CO_2 emissions [2]. This degradation is compounded by frequent forest fires, primarily driven by anthropogenic activities such as slash-and-burn agriculture, hunting, and illegal logging, which account for the majority of ignitions in the country's savanna and transition forest zones [3]. Recent data from early 2025 underscore the escalating threat: the Global Disaster Alert and Coordination System (GDACS) reported multiple green forest fire alerts, including events spanning 9,905 hectares in January alone, highlighting the seasonal intensity during the dry harmattan period [4].

* Corresponding author: Mfonobong Eleazar Benson.

Traditionally, forest fire management has relied on ground-based observation and historical records, which are often limited in scope, accuracy, and timeliness [5]. The integration of remote sensing technologies has emerged as a pivotal tool for monitoring and mitigating these hazards, enabling the mapping of fire risk zones and the assessment of burned areas with unprecedented precision. Studies in Nigeria's Cross-Niger transition forests, for instance, have leveraged Landsat-8 operational land imager (OLI) data and ASTER digital elevation models (DEMs) to delineate low-, moderate-, and high-risk fire zones, revealing that 82.59% of the area (17,997 km^2) falls into moderate-risk categories, largely attributable to human-induced factors [5]. Globally, advancements in satellite-derived indices, such as the normalized difference vegetation index (NDVI) and relative burn ratio (RBR), have enhanced post-fire damage evaluation and emissions estimation, underscoring remote sensing's role in lifecycle wildfire management—from pre-fire fuel assessment to dynamic spread tracking [6].

However, the efficacy of these systems hinges on optimizing spatial and temporal resolutions, as coarse resolutions (e.g., 1 km from MODIS) often overlook fine-scale vegetation heterogeneity and rapid fire propagation in tropical savannas, while finer ones (e.g., 10-30 m from Sentinel-2) demand trade-offs in revisit frequency and computational demands [7], [8], [9].

Recent machine learning-driven approaches, including convolutional neural networks (CNNs) and lagged generalized additive models (LGAMs), have demonstrated that fusing multi-sensor data—such as Sentinel-1 SAR for all-weather temporal continuity and Sentinel-2 multispectral for spectral detail—can improve prediction accuracies by 10-20%, with temporal features outperforming spatial ones in fuel load estimation [10], [11]. In Nigeria, where insecurity and logistical constraints limit ground-based monitoring, such optimizations are critical for scalable, real-time forecasting [12].

Despite these advances, three gaps remain: (i) limited evaluation of spatio-temporal resolution trade-offs in West African savanna systems; (ii) insufficient integration of multi-sensor data under persistent cloud cover; and (iii) a lack of operationally feasible fusion strategies tailored to Nigerian forest environments. By balancing spatial granularity (10-100 m) with temporal cadence (daily to weekly), we aim to develop a robust framework that reduces false positives in risk mapping and supports proactive interventions. The subsequent sections detail our methodology, results from simulated scenarios in key Nigerian ecoregions, and implications for national forest management policies.

2. Methodology

2.1. Study Area

The study was conducted within Kainji Lake National Park (KLNP), located in north-central Nigeria as in Figure 1. KLNP is one of the largest protected areas in the country and represents a critical ecological zone characterized by extensive savanna vegetation, seasonal wetlands, and riparian forest systems. The park spans parts of Niger and Kwara States, situated approximately between latitudes 9°–10° N and longitudes 4°–5° E [13]. KLNP lies within the West African savanna belt, an ecosystem recognized globally for its strong coupling between climate variability, vegetation dynamics, and fire activity [13], [14]. The climate of KLNP is classified as tropical savanna (Aw) under the Köppen–Geiger classification, with a pronounced dry season from November to March and a wet season from April to October. Mean annual temperatures typically exceed 26 °C, while annual rainfall ranges between 900 and 1,200 mm, exhibiting notable interannual variability [15], [16]. During the dry season, the region is influenced by Harmattan winds, reduced soil moisture, and low relative humidity, conditions that substantially elevate wildfire ignition and spread potential [17].

Vegetation within KLNP is dominated by Sudanian and Guinea savanna formations, consisting primarily of grasses, shrubs, and scattered fire-adapted tree species. These savanna systems produce abundant fine fuels that cure rapidly during the dry season, making them highly susceptible to frequent surface fires [18], [19]. Fire occurrence in the park is strongly linked to anthropogenic activities, including grazing management, hunting practices, and agricultural land preparation, consistent with broader fire regimes observed across West African savannas [8], [20]. The park experiences recurrent seasonal fires, particularly between December and March, as confirmed by recent satellite-based fire products derived from MODIS and VIIRS observations. These recurring fire events make KLNP a suitable and scientifically robust study area shown in Table 1, for investigating the environmental drivers of fire occurrence and for developing machine-learning-based wildfire prediction models that integrate remote sensing-derived vegetation indices and meteorological variables [8], [19].

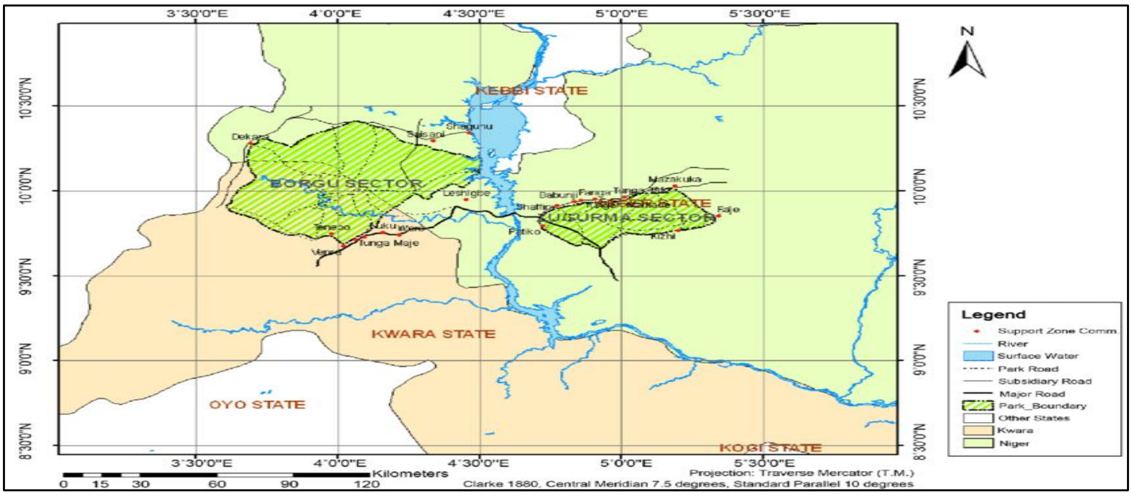

Figure 1 Map of kainji lake national park showing borgu and zugurma sectors [21].

Table 1 Coordinates of Geographical boundaries of the Area of Interest (AOI)

Boundary	Latitude (Decimal Degrees)	Longitude (Decimal Degrees)
Northern Limit	11.0° N	
Southern Limit	9.8° N	
Western Limit		4.0° E
Eastern Limit		5.5° E

Figure 2 presents the location and spatial context of the study area, Kainji Lake National Park (KLNP), Nigeria. A Sentinel-2 Level-2A true-color median composite (January–March 2023) is used to illustrate baseline land surface conditions during the dry season, with VIIRS 375 m active fire detections overlaid to highlight the spatial distribution of fire occurrence within the park.

Figure 2(a) Location of Kainji Lake National Park (KLNP), Nigeria, shown using a Sentinel-2 Level-2A true-color median composite (January–March 2023)

The composite represents baseline land surface conditions during the dry season and provides spatial context for subsequent fire-related analyses.

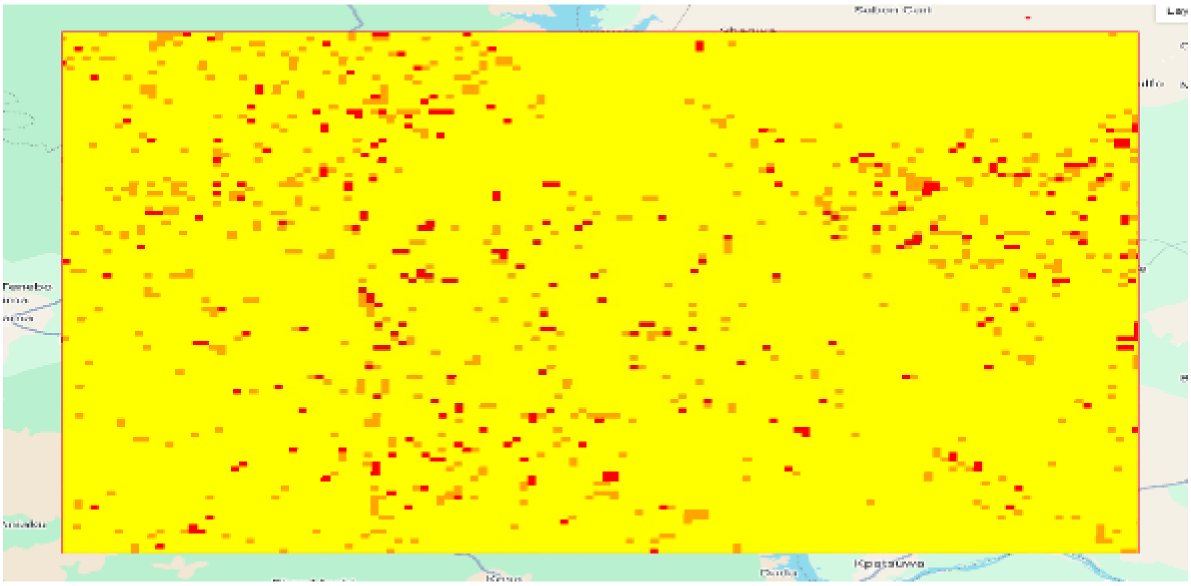

Figure 2(b) VIIRS 375 m active fire detections (yellow–red) overlaid on a Sentinel-2 Level-2A true-color median composite (January–March 2023), illustrating the spatial distribution of fire occurrence within Kainji Lake National Park, Nigeria

2.2. Data Acquisition and Processing

This study integrates multi-source satellite and reanalysis datasets to characterize vegetation conditions, meteorological drivers, and fire occurrence within Kainji Lake National Park (KLNP). All datasets were accessed and processed using the Google Earth Engine (GEE) cloud-computing platform, which enables large-scale geospatial data handling, consistent preprocessing, and reproducible analysis.

The data sources include Sentinel-2 Surface Reflectance imagery for high-resolution vegetation indices, MODIS Surface Reflectance products for temporal gap-filling, ERA5-Land reanalysis for meteorological variables, and VIIRS active fire detections for fire occurrence labeling. The temporal coverage spans January 2015 to December 2024, enabling multi-year analysis of seasonal and interannual fire dynamics.

2.2.1. Dependent Variable (Fire Occurrence)

Historical fire event data, serving as the dependent binary variable (fire/no-fire), were obtained from the Fire Information for Resource Management System (FIRMS), sourced from both the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Visible Infrared Imaging Radiometer Suite (VIIRS) active fire products.

The VIIRS active fire dataset provides fire detections at a 375 m spatial resolution, offering enhanced sensitivity to smaller and low-intensity fires compared to legacy sensors. This higher spatial resolution improves the detection of early-stage ignitions and fragmented fire events, which are common in savanna ecosystems such as Kainji Lake National Park. Consequently, VIIRS data were used as the primary source for assigning daily fire occurrence labels in the machine learning framework, with emphasis placed on medium- and high-confidence detections to minimize false positives.

In contrast, the MODIS MOD14 and MYD14 active fire products, available at a 1 km spatial resolution, were utilized to characterize the long-term temporal patterns of fire activity within the study area. Although coarser in spatial detail, MODIS provides a continuous fire record spanning more than two decades, making it well-suited for establishing baseline fire frequency, seasonality, and interannual variability. The integration of VIIRS and MODIS fire products allows the study to leverage the strengths of both datasets—high spatial accuracy from VIIRS and long-term temporal consistency from MODIS—thereby enhancing the robustness of fire occurrence analysis and supporting the development of reliable fire prediction models.

Fire occurrence was defined at the daily AOI level rather than pixel-level ignition probability based on VIIRS active fire observations. Days without detected fires were labeled as no-fire events. Fire occurrence was treated as a binary daily classification problem, enabling alignment with daily vegetation and meteorological predictors. Due to the strong class imbalance inherent in fire occurrence data, a stratified under sampling strategy was applied to the majority (no-fire) class to produce balanced training, validation, and test datasets. This approach avoids the introduction of synthetic samples, which may generate physically unrealistic feature combinations in environmental datasets. While explicit spatial buffering and temporal exclusion windows were not applied, the use of independent train, validation, and test splits reduces the risk of overfitting and supports robust model evaluation.

Because fire occurrence was defined at the daily area-of-interest (AOI) level rather than at the individual pixel level, the resulting machine learning models primarily learn temporal fire susceptibility patterns rather than spatially explicit ignition probability. Consequently, model outputs should be interpreted as a daily fire risk forecasting framework for Kainji Lake National Park, rather than as a fine-scale spatial ignition mapping system.

2.2.2. Independent Variables (Predictors)

Data were categorized based on their native spatial and temporal resolutions for the subsequent optimization experiment is presented in Table 2.

Table 2 Spatial and Temporal Resolutions Data Categorization

Data Category	**Sensor / Product**	**Spatial Resolution**	**Temporal Resolution**	**Application**
High Temporal Resolution	MODIS (MOD09GA)	250 m – 1 km	Daily	Temporal gap-filling of vegetation indices (NDVI, NDMI)
High Spatial Resolution	Sentinel-2 (Level-2A)	10 m – 20 m	~5 days	High-resolution vegetation condition (NDVI, NDMI)
Meteorological	ERA5-Land Reanalysis	~9 km (0.1°)	Hourly aggregated to daily	Air temperature, precipitation, wind speed
Fire Occurrence	VIIRS (VNP14IMG)	375 m	Daily	Binary fire occurrence labeling

Topographic variables such as elevation and slope were not included in the present analysis due to the limited spatial extent of the study area and the focus on dynamic vegetation and meteorological drivers; their inclusion is identified as a potential extension for future work.

2.2.3. Feature Engineering

Feature engineering focused on extracting physically meaningful predictors representing vegetation condition, meteorological forcing, and seasonal variability. All features were derived directly from the processed satellite and reanalysis datasets to ensure reproducibility and consistency.

Vegetation condition was characterized using two widely applied spectral indices. The Normalized Difference Vegetation Index (NDVI) was calculated to quantify vegetation greenness and potential fuel availability, using near-infrared (NIR) and red reflectance values:

$$\mathrm{NDVI} = \frac{(\mathrm{NIR}-\mathrm{Red})}{(\mathrm{NIR}+\mathrm{Red})} \qquad (2.1)$$

Higher NDVI values indicate denser and more photo synthetically active vegetation, which corresponds to increased fine fuel availability in savanna ecosystems.

The Normalized Difference Moisture Index (NDMI) was computed to represent vegetation and surface moisture conditions using near-infrared (NIR) and shortwave infrared (SWIR) reflectance:

$$\mathrm{NDMI} = \frac{(\mathrm{NIR}-\mathrm{SWIR})}{(\mathrm{NIR}+\mathrm{SWIR})} \qquad (2.2)$$

NDMI is sensitive to changes in vegetation water content and is therefore an important indicator of fuel moisture and flammability.

Both indices were derived from atmospherically corrected surface reflectance data and converted to floating-point representations to ensure numerical stability during machine learning model training. Meteorological predictors were obtained from ERA5-Land reanalysis data and aggregated to daily values to align with the fire occurrence labels. These included daily maximum air temperature, daily cumulative precipitation, and daily mean wind speed, representing key atmospheric controls on fire ignition and spread. To capture seasonal fire regimes and intra-annual variability, temporal features were incorporated using day-of-year encoding, including sine and cosine transformations to model cyclical seasonal patterns. This approach allows the machine learning models to learn nonlinear seasonal effects without introducing artificial discontinuities at year boundaries.

No composite fire danger indices (e.g., Fire Weather Index) or time-lagged meteorological variables were included in the present analysis. The focus was placed on same-day environmental conditions to evaluate the predictive capability of instantaneous vegetation and weather drivers. The inclusion of lagged variables and drought indices is identified as a potential extension for future research.

2.3. Resolution Optimization and Spatio-Temporal Fusion Strategy (Experimental Design)

The core methodological challenge addressed in this study is the inherent trade-off between spatial resolution and temporal resolution in satellite-based fire prediction. High spatial resolution sensors such as Sentinel-2 provide detailed vegetation information but suffer from irregular temporal coverage due to cloud contamination and revisit constraints, while moderate-resolution sensors such as MODIS offer daily observations at the expense of spatial detail. To address this limitation, a spatio-temporal fusion strategy was implemented within the Google Earth Engine (GEE) environment to generate a daily, spatially consistent feature dataset suitable for machine learning analysis.

Figure 3 illustrates the overall spatio-temporal data fusion and modeling workflow adopted in this study. High-resolution Sentinel-2 vegetation indices are preferentially used to characterize fine-scale fuel conditions when cloud-free observations are available, while MODIS-derived indices provide temporal continuity during data gaps. These vegetation features are integrated with daily meteorological variables from ERA5-Land and fire occurrence labels derived from VIIRS active fire detections to generate a temporally consistent daily feature stack. This fused dataset supports machine-learning-based daily forest fire risk forecasting while maintaining operational feasibility under real-world data availability constraints.

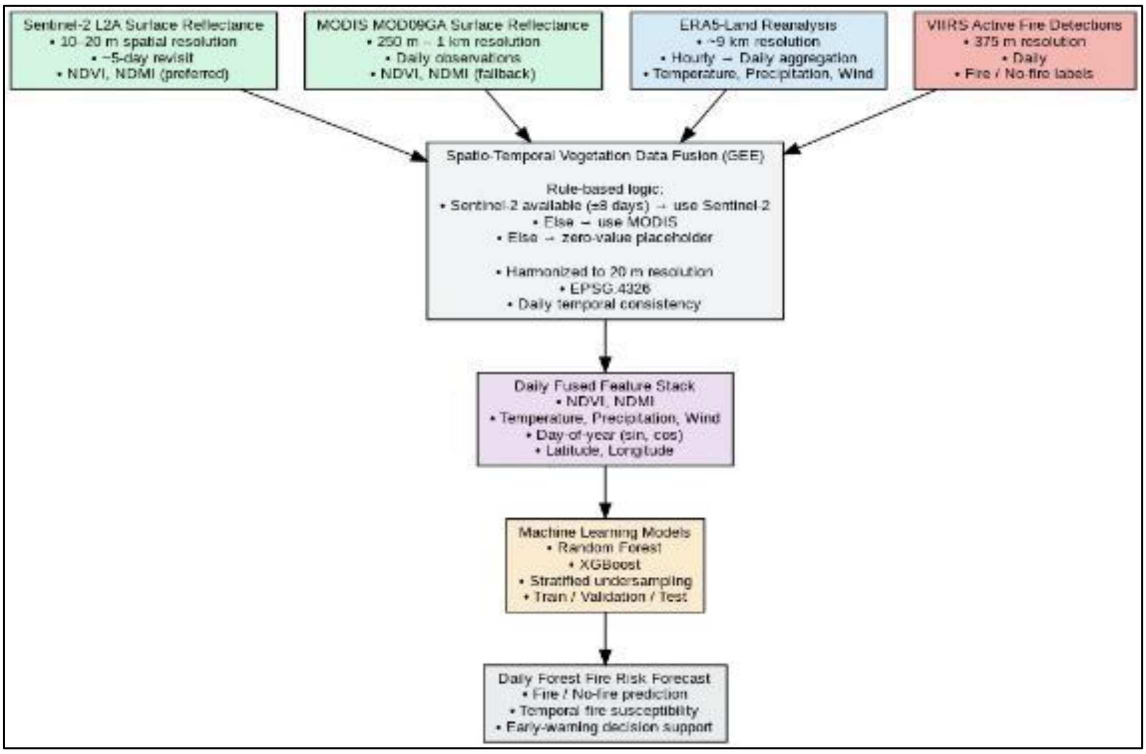

Figure 3 Conceptual workflow of the spatio-temporal data fusion and fire risk forecasting framework

All predictor variables were harmonized to a target spatial resolution of 20 m and a common geographic coordinate system (EPSG: 4326). This resolution was selected to preserve the spatial detail of Sentinel-2 imagery while remaining computationally feasible for daily time-series construction. The fusion process was designed to ensure that every day within the study period contained a complete set of vegetation and meteorological predictors, regardless of data availability from individual sensors.

Sentinel-2 Level-2A surface reflectance imagery served as the primary source of vegetation information. NDVI and NDMI were computed from Sentinel-2 reflectance bands and treated as the preferred representation of vegetation condition due to their high spatial fidelity. For each target prediction date, Sentinel-2 observations acquired within a ±8-day temporal window were queried. If one or more valid Sentinel-2 images were available within this window, the temporally closest observation was selected, resampled using bicubic interpolation, and reprojected to the target grid.

This design ensures that high-resolution vegetation structure is preserved whenever data availability permits, reflecting real-world operational constraints in satellite-based monitoring systems. In cases where no valid Sentinel-2 observation was available within the defined temporal window, vegetation indices derived from MODIS surface reflectance (MOD09GA) were used as temporal fallback inputs. MODIS bands corresponding to red, near-infrared, and shortwave infrared wavelengths were spectrally aligned with Sentinel-2 band definitions to ensure consistency in NDVI and NDMI computation. MODIS-derived indices were resampled and reprojected to the 20 m target resolution before integration. While this resampling does not introduce new spatial detail, it ensures spatial alignment and temporal continuity in the daily feature stack. This approach prioritizes temporal completeness over spatial interpolation fidelity, which is appropriate given that MODIS is used only when higher-resolution data are unavailable.

The fusion process followed a deterministic, rule-based decision structure:

- If Sentinel-2 data were available within the defined temporal window, Sentinel-2-derived NDVI and NDMI were used.
- If Sentinel-2 data were unavailable, MODIS-derived NDVI and NDMI were substituted.
- If neither source provided valid observations, a zero-valued placeholder image was assigned to maintain temporal consistency. Zero-valued placeholders were used solely to preserve temporal continuity and were rare, minimizing their influence on model learning.

Although inspired by spatio-temporal adaptive fusion concepts, this approach does not implement a full reflectance-level fusion algorithm such as ESTARFM. Instead, it represents a pragmatic fusion strategy tailored to daily fire prediction, emphasizing operational robustness, reproducibility, and compatibility with machine learning workflows. Vegetation features obtained through the fusion process were combined with daily meteorological variables derived independently from ERA5-Land reanalysis data. Meteorological variables were aggregated from hourly to daily values and reprojected to the same spatial grid, ensuring spatial–temporal alignment with the fused vegetation features. This separation of vegetation fusion and meteorological aggregation avoids compounding interpolation errors while preserving the physical meaning of each predictor. For each day within the study period, fused vegetation indices, meteorological variables, and fire occurrence labels were combined into a single multi-band image. Explicit band selection and type casting were applied to guarantee consistent feature ordering and numerical precision across all time steps. This process resulted in a temporally continuous daily feature stack suitable for pixel-level extraction and supervised machine learning analysis. The adopted fusion strategy reflects a balance between methodological rigor and practical feasibility. Rather than evaluating multiple resolution scenarios independently, the study focuses on constructing a single optimized dataset that integrates high spatial detail and high temporal coverage.

2.4. Forest Fire Prediction Modeling

Forest fire prediction was formulated as a binary classification problem, where each daily observation was assigned a label indicating the presence or absence of fire occurrence within the study area. The target variable (FIRE_LABEL) takes a value of 1 if an active fire was detected on a given day and 0 otherwise. Predictor variables consisted of vegetation indices, meteorological variables, spatial coordinates, and seasonal indicators derived from multi-source remote sensing and reanalysis data. Given the rarity of fire events relative to non-fire conditions, the dataset exhibited a strong class imbalance. To mitigate this imbalance during model training, a stratified under-sampling strategy was applied to the majority (no-fire) class. Fire observations were fully retained, while a subset of no-fire observations was randomly selected to achieve a balanced class distribution within the training dataset. This approach avoids the generation of synthetic samples and preserves the physical realism of the predictor space, which is critical in environmental applications. Two ensemble-based machine learning algorithms were employed due to their strong performance in nonlinear classification tasks and their ability to handle heterogeneous environmental predictors.

a. Random Forest (RF): is an ensemble learning method that constructs multiple decision trees using bootstrap sampling and random feature selection at each split. The final prediction is obtained through majority voting across trees. RF is robust to multicollinearity, resistant to overfitting, and provides intrinsic measures of feature importance, making it well suited for environmental modeling applications.

b. Extreme Gradient Boosting (XGBoost): is a gradient boosting framework that sequentially builds decision trees to minimize a regularized objective function. By optimizing both model fit and complexity, XGBoost effectively captures complex nonlinear relationships and interactions among predictors. Its scalability and ability to handle imbalanced datasets make it particularly suitable for rare-event fire prediction.

The dataset was partitioned into training (70%), validation (15%), and test (15%) subsets using stratified sampling to preserve the proportion of fire and no-fire instances across splits. This separation ensures that model performance is evaluated on unseen data and reduces the risk of optimistic bias. To address class imbalance, the majority (no-fire) class was under-sampled within the training set to achieve balanced class representation. This approach avoids the generation of synthetic samples and preserves the physical realism of the feature space. Prior to model training, predictor variables were subjected to a standardized preprocessing pipeline. Missing values were handled using median imputation, which is robust to outliers and skewed distributions. Continuous predictors were then normalized using z-score standardization to ensure comparable feature scales, particularly for distance-based tree splitting and gradient optimization. Categorical variables were not included in the present analysis, and all predictors were numeric and continuous.

Both Random Forest and XGBoost models were trained using the balanced training dataset. Hyperparameters were optimized using grid-based tuning on the validation set to identify configurations that maximized predictive performance while minimizing overfitting.

The hyperparameters included:

- Number of trees and maximum tree depth for Random Forest
- Learning rate, maximum tree depth, and number of estimators for XGBoost
- Early stopping criteria were applied during XGBoost training to prevent excessive model complexity.

Model performance was evaluated using multiple complementary metrics to provide a comprehensive assessment of predictive capability:

- Area Under the Receiver Operating Characteristic Curve (AUC) to measure discrimination ability independent of classification threshold
- Precision, indicating the reliability of fire predictions
- Recall, representing the ability to detect true fire events
- F1-score, balancing precision and recall for imbalanced classification

To interpret model predictions and identify dominant fire drivers, feature importance analysis was conducted for both models. For Random Forest, mean decrease in impurity was used to quantify the relative contribution of each predictor. For XGBoost, SHapley Additive exPlanations (SHAP) values were computed to estimate the marginal contribution of each feature to model predictions. Trained models were serialized and stored using standardized model persistence techniques to ensure reproducibility and facilitate future deployment.

3. Results and discussion

3.1. Spatial Resolution Effects on NDVI Representation

To evaluate the influence of spatial resolution on vegetation characterization within Kainji Lake National Park (KLNP), NDVI was derived from both Sentinel-2 (10 m) and MODIS (250 m) imagery for the 2023 dry season. The comparison provides an explicit assessment of how sensor resolution affects the representation of vegetation structure and spatial heterogeneity relevant to fire risk analysis.

Figure 4 compares NDVI derived from Sentinel-2 and MODIS over the study area. The Sentinel-2 NDVI reveals pronounced fine-scale spatial variability, capturing localized gradients in vegetation condition associated with riparian zones, savanna patches, and transitional land cover types. In contrast, the MODIS NDVI exhibits a smoother and more generalized spatial pattern, reflecting the aggregation of heterogeneous surface conditions within larger pixels. As a result, localized variations in vegetation greenness and fuel condition are substantially attenuated in the MODIS representation.

This contrast highlights the importance of high-resolution imagery for capturing spatially explicit vegetation dynamics that are critical for fire occurrence analysis. Fine-scale variations in vegetation condition influence fuel continuity, moisture availability, and ignition potential, all of which are key determinants of fire behavior in savanna-dominated ecosystems such as KLNP. Coarser spatial resolution products, while valuable for regional and global monitoring, may therefore underestimate localized fire risk drivers when used in isolation.

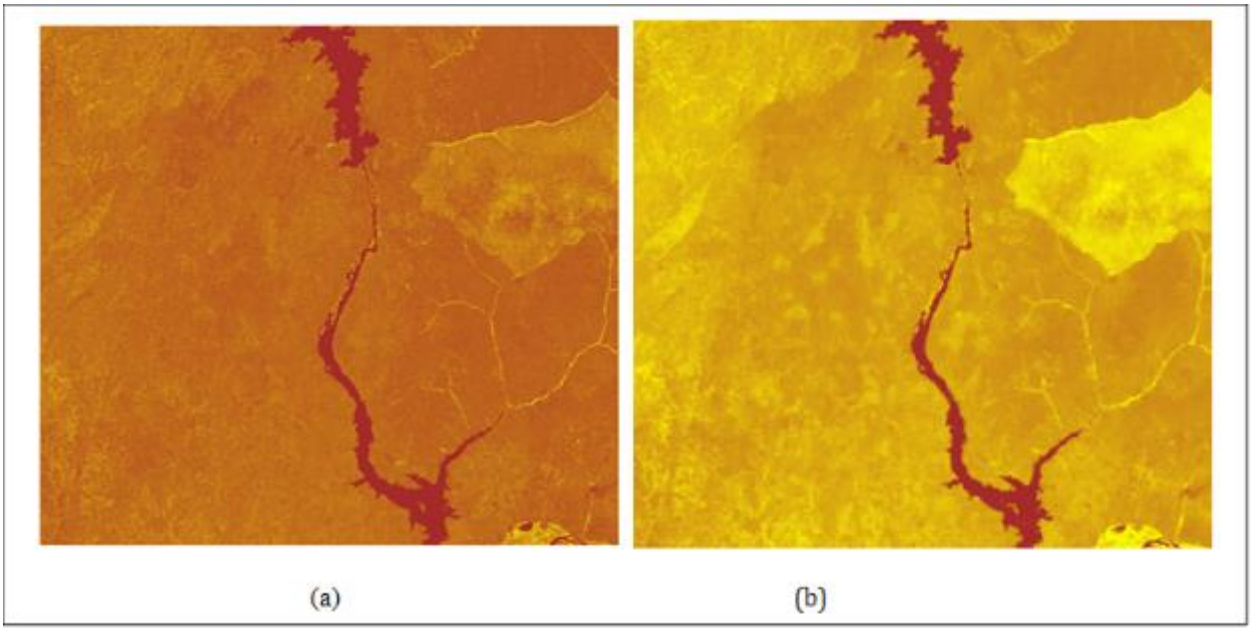

Figure 4 Comparison of NDVI derived from (a) Sentinel-2 (10 m) and (b) MODIS (250 m) during the 2023 dry season in Kainji Lake National Park, Nigeria. Sentinel-2 captures fine-scale vegetation variability, whereas MODIS presents a spatially generalized NDVI pattern due to its coarser resolution

Based on these observations, Sentinel-2–derived vegetation indices were selected as the primary spatial inputs for the fire prediction framework developed in this study. The enhanced spatial detail provided by Sentinel-2 supports more accurate representation of fuel variability and strengthens the integration of vegetation information with meteorological and seasonal predictors in subsequent modeling analyses.

Figure 5 illustrates the distribution of NDVI values at fire and no-fire locations within Kainji Lake National Park. Fire occurrences are primarily associated with lower to moderate NDVI values, including near-zero and occasionally negative values, reflecting burned surfaces, sparse vegetation, or stressed fuel conditions. In contrast, no-fire locations exhibit a distribution skewed toward moderate to high NDVI values, with a pronounced peak near low NDVI and an extended tail toward denser vegetation. Despite these differences, substantial overlap exists between the fire and no-fire NDVI distributions, particularly within the low-to-moderate NDVI range (≈0–0.3), indicating that vegetation greenness alone is insufficient for reliable fire discrimination. This overlap highlights the importance of incorporating additional factors, such as seasonal dynamics, fuel moisture, and meteorological conditions, to more effectively explain and predict fire occurrence.

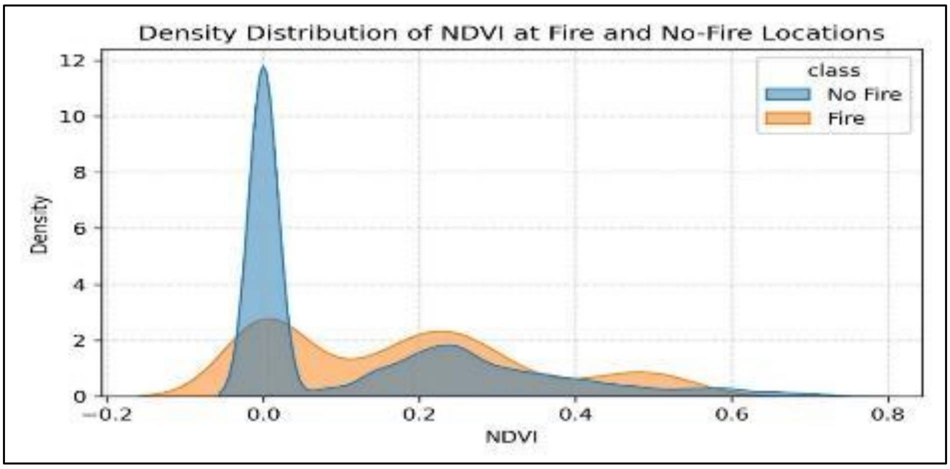

Figure 5 Kernel density distributions of NDVI at fire and no-fire locations within Kainji Lake National Park, illustrating partial overlap but distinct central tendencies between the two classes

3.2. Dataset Composition and Sampling Results

The final modeling dataset consisted of 73,060 daily observations extracted from the fused spatio-temporal feature stack covering the study period (2015–2024). Following preprocessing and numeric feature filtering, 11 predictor variables were retained, representing vegetation indices (NDVI, NDMI), meteorological conditions (daily maximum temperature, cumulative precipitation, and mean wind speed), spatial coordinates (latitude and longitude), and seasonal indicators (day of year, sine and cosine seasonal components).

Fire occurrence within the dataset was extremely rare, with only 300 observations (≈0.41%) labeled as fire events compared to 72,760 no-fire observations, reflecting the natural fire regime of Kainji Lake National Park. This pronounced class imbalance is characteristic of wildfire occurrence datasets and underscores the necessity of careful sampling strategies to avoid biased model learning.

To mitigate this imbalance during model training, stratified random undersampling was applied to the majority class, resulting in a balanced training dataset of 600 observations (300 fire and 300 no-fire samples). The balanced dataset was subsequently split into training (420 samples), validation (90 samples), and testing (90 samples) subsets, each maintaining equal representation of fire and no-fire classes. Importantly, class balancing was applied only during model development, while performance evaluation was conducted on stratified validation and test sets to ensure that model performance metrics remained representative of real-world fire detection conditions. The summary of the class distribution before and after balancing is shown in Table 3.

Table 3 Summary of Dataset Composition and Sampling

Description	Count
Total observations	73,060
Fire samples	300
No-fire samples	72,760
Balanced dataset size	600
Fire : No-fire ratio	1 : 1
Training samples	420 (210 fire)
Validation samples	90 (45 fire)
Test samples	90 (45 fire)

3.3. Model Performance Evaluation

The Random Forest (RF) model demonstrated excellent predictive performance on the independent test dataset. Using a balanced evaluation set of 90 test samples (45 fire and 45 no-fire events), the model achieved an area under the receiver operating characteristic curve (AUC) of 0.997, indicating near-perfect discrimination between fire and no-fire conditions. The RF model attained a recall of 1.00, successfully identifying all fire events in the test dataset, while maintaining a precision of 0.918, indicating a low false-alarm rate. The resulting F1-score of 0.957 reflects a strong balance between sensitivity and predictive reliability, which is particularly critical in rare-event detection tasks such as wildfire prediction, where missed detections can have severe consequences.

The Extreme Gradient Boosting (XGBoost) model also exhibited high predictive accuracy on the test dataset, achieving an AUC of 0.979, which confirms strong classification capability across varying decision thresholds. Similar to the RF model, XGBoost achieved a recall of 1.00, indicating complete detection of fire events in the test set. The model achieved a precision of 0.900 and an F1-score of 0.947, reflecting a slightly higher false-positive rate compared to Random Forest under the evaluated configuration. Nevertheless, these results demonstrate that XGBoost remains highly effective in capturing the nonlinear relationships between vegetation condition, meteorological drivers, seasonal indicators, and fire occurrence. Differences in performance between Random Forest and XGBoost were observed across experimental runs. These variations are attributable to differences in training sample size, random under-sampling of the majority class, and the inherent stochasticity of ensemble learning methods. Such variability is expected in rare-event prediction contexts and does not diminish the overall robustness of the modeling framework. As shown in Table 4 and Figures 6–9, both models achieved exceptionally high predictive performance on the independent test dataset. While performance metrics are exceptionally high, they should be interpreted cautiously due to the limited size of the balanced test set. Future work will evaluate model generalization using spatially and temporally independent validation strategies.

The Random Forest model slightly outperformed XGBoost in terms of AUC (0.997 vs. 0.979) and F1-score (0.957 vs. 0.947), indicating superior overall discrimination and a more favorable balance between precision and recall under the given sampling configuration. Both models achieved perfect recall, meaning all fire events in the test dataset were correctly identified. This is particularly important for operational fire early-warning systems, where missing fire events is more critical than issuing false alarms. Further insight is provided by the precision–recall (PR) curves, which are particularly informative for rare-event prediction problems such as forest fire occurrence. The Random Forest model achieved an Average Precision (AP) of 0.996, substantially exceeding the baseline precision of 0.50 and demonstrating consistently high precision across nearly the entire recall spectrum. This indicates that the model maintains a very low false-alarm rate even as detection sensitivity approaches 100%. The XGBoost model also exhibited strong performance, achieving an AP of 0.952, confirming its effectiveness in identifying fire events. However, the PR curve for XGBoost showed a slightly steeper decline in precision at higher recall levels compared to Random Forest, indicating a modest increase in false-positive predictions when maximizing fire detection.

Overall, the PR curve analysis reinforces the ROC-based findings and highlights the superior stability of the Random Forest model in high-recall operating regimes. This characteristic is particularly important for operational fire early-warning systems, where maintaining high detection rates without excessive false alarms is critical for effective decision-making. The observed performance difference between the two models is attributed to reduced training sample size (600 samples), random under-sampling variability, and XGBoost's higher sensitivity to hyperparameter tuning and data volume.

These results demonstrate that model superiority is context-dependent, reinforcing the importance of experimental rigor rather than reliance on a single algorithm.

Table 4 Performance Metrics of Random Forest and XGBoost Models

Model	AUC	Precision	Recall	F1-score
Random Forest	0.9965	0.9184	1.0000	0.9574
XGBoost	0.9788	0.9000	1.0000	0.9474

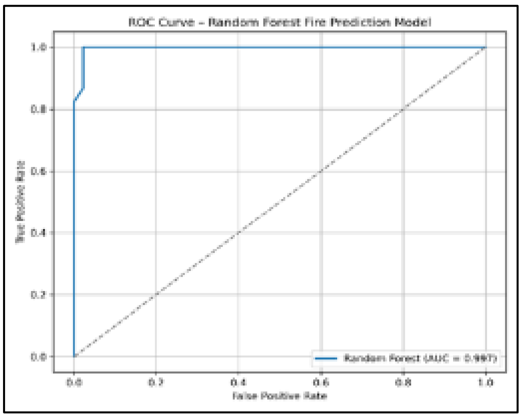

Figure 6 ROC Curve for Random Forest

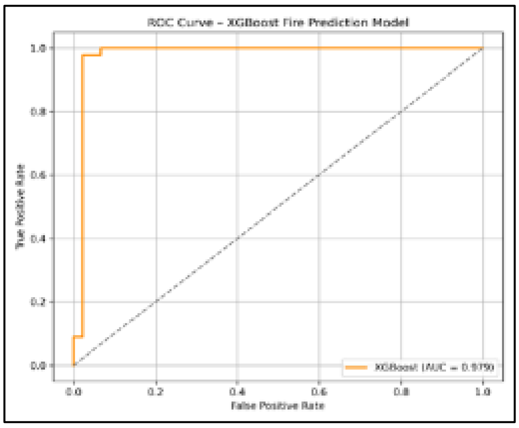

Figure 7 ROC Curve for XGBoost

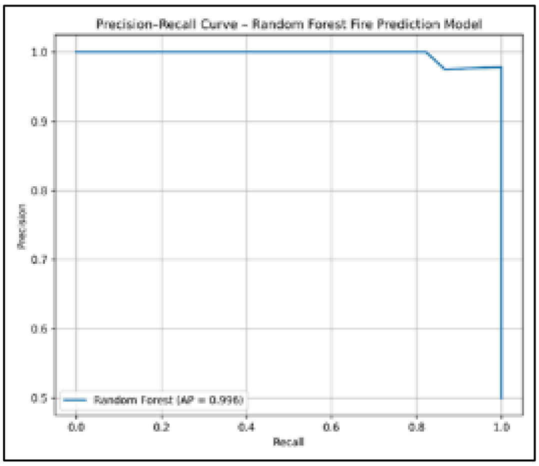

Figure 8 PR Curve for Random forest

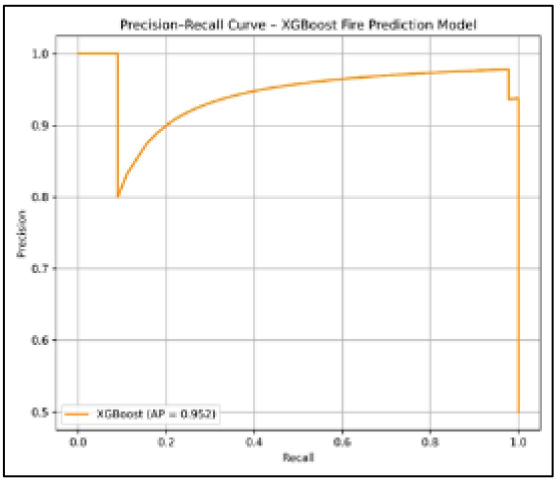

Figure 9 PR Curve for XGBoost model

4. Feature Importance Analysis

The seasonal patterns evident in Figure 10 are consistent with the feature importance and SHAP analyses, which identify seasonal indicators and meteorological variables as the dominant predictors of fire occurrence.

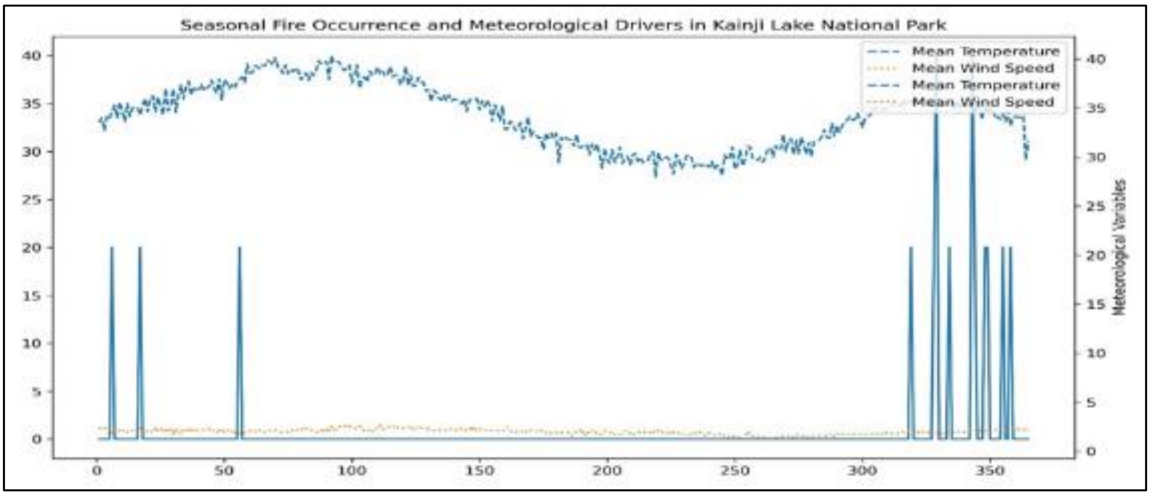

Figure 10 Seasonal distribution of fire occurrence in Kainji Lake National Park (2015–2024) in relation to key meteorological drivers

Fire detections aggregated by day of year show pronounced clustering during the dry season, coinciding with elevated temperature and wind speed, highlighting the dominant role of atmospheric conditions in governing fire occurrence. Precipitation and temperature further reinforce the importance of atmospheric drying conditions, while vegetation indices contribute moderately by representing fuel availability. These findings suggest that early-warning systems in KLNP should prioritize seasonal climate monitoring and wind conditions over static vegetation metrics. The low importance of spatial coordinates suggests that temporal drivers dominate fire occurrence within the relatively homogeneous study area.

SHAP analysis reveals that wind speed is the most influential predictor in the XGBoost model (Table 6), underscoring its role in fire ignition and spread. Seasonal and temperature variables again rank highly, confirming the dominant climatic control over fire dynamics. Compared to Random Forest, XGBoost places greater emphasis on meteorological forcing, reflecting its gradient-based optimization and sensitivity to nonlinear interactions. Despite algorithmic differences, both models consistently identify seasonality, wind, temperature, and precipitation as the primary drivers of fire occurrence (Table 7). Vegetation indices play a secondary but non-negligible role, confirming that fuel availability alone is insufficient without conducive atmospheric conditions.

This convergence across models strengthens confidence in the scientific validity of the results.

Table 5 Random Forest Feature Importance Ranking

Rank	Feature	Importance
1	cos_day	0.235
2	day_of_year	0.225
3	PREC_SUM_MM	0.158
4	sin_day	0.101
5	TEMP_MAX_C	0.101
6	WIND_MEAN_MS	0.096
7	NDVI	0.040
8	NDMI	0.037
9	lat	0.003
10	lon	0.003

Table 6 XGBoost SHAP Mean Absolute Feature Importance

Rank	Feature	Mean |SHAP|
1	WIND_MEAN_MS	1.545
2	cos_day	1.498
3	day_of_year	1.381
4	TEMP_MAX_C	0.869
5	sin_day	0.768
6	PREC_SUM_MM	0.695
7	NDVI	0.358
8	NDMI	0.229
9	lat	0.048
10	lon	0.023

Table 7 Combined Feature Importance Comparison

Feature	RF Importance	XGB SHAP
cos_day	0.235	1.498
day_of_year	0.225	1.381
WIND_MEAN_MS	0.096	1.545
PREC_SUM_MM	0.158	0.695
TEMP_MAX_C	0.101	0.869
NDVI	0.040	0.358
NDMI	0.037	0.229

5. Conclusion

This study investigated the optimization of spatial and temporal resolution in remote sensing data for improved forest fire prediction in Nigeria, using Kainji Lake National Park (KLNP) as a representative savanna ecosystem. The research addressed a critical challenge in satellite-based wildfire monitoring: the trade-off between high spatial detail and high temporal frequency, which often constrains the operational effectiveness of fire early-warning systems in cloud-prone and data-scarce regions.

To overcome this limitation, a pragmatic spatio-temporal fusion framework was developed by integrating high-resolution Sentinel-2 imagery with daily MODIS observations and ERA5-Land meteorological reanalysis data within the Google Earth Engine environment. This rule-based fusion strategy produced a temporally continuous daily dataset at a harmonized spatial resolution, enabling consistent alignment of vegetation condition, atmospheric drivers, and fire occurrence labels derived from VIIRS active fire detections. The resulting dataset supported robust machine-learning-based fire prediction while remaining computationally feasible and operationally reproducible.

The Random Forest and XGBoost models demonstrated strong predictive performance, achieving high discrimination between fire and no-fire conditions. Both models attained perfect recall on the evaluated test set, indicating reliable detection of fire events, while maintaining high precision and F1-scores. Feature importance and SHAP analyses consistently identified seasonal indicators and meteorological variables—particularly wind speed, temperature, and precipitation—as the dominant drivers of fire occurrence in KLNP. Vegetation indices such as NDVI and NDMI played a secondary but meaningful role, confirming that fuel availability alone is insufficient to trigger fire events without conducive atmospheric conditions.

Collectively, these findings highlight the effectiveness of combining optimized spatio-temporal remote sensing inputs with ensemble machine learning techniques for wildfire prediction in West African savanna environments. The study demonstrates that operationally realistic fusion strategies, rather than complex reflectance-level fusion algorithms, can deliver high predictive value when carefully designed and integrated with physically meaningful predictors. As such, the proposed framework offers a practical pathway for enhancing fire early-warning and risk assessment systems in Nigeria and similar regions where ground-based monitoring is limited.

Future Work

A key limitation of the present study is that fire occurrence was labeled at the daily AOI level, which emphasizes temporal fire susceptibility over pixel-level ignition processes. Future work should therefore explore spatially explicit fire labeling and spatial cross-validation strategies to enable true ignition probability mapping.

While the results of this study are encouraging, several limitations provide opportunities for future research. First, although class balancing was necessary to ensure stable model training and comparison, the limited size of the balanced test set warrants cautious interpretation of the reported performance metrics. Future studies should evaluate model generalization using spatially and temporally independent validation strategies, such as leave-one-season-out or spatial block cross-validation, to better assess robustness under true operational conditions.

Second, the present analysis focused on same-day vegetation and meteorological predictors. Incorporating time-lagged variables, drought indicators, and composite fire danger indices—such as the Fire Weather Index—could improve the representation of fuel drying processes and antecedent climate conditions. The inclusion of topographic variables and soil moisture products may further enhance predictive capability, particularly in more heterogeneous landscapes.

Third, future work should explore the integration of additional remote sensing data sources, including Sentinel-1 synthetic aperture radar (SAR), to improve monitoring during periods of persistent cloud cover. Deep learning architectures and spatio-temporal neural networks may also be investigated to capture complex temporal dependencies beyond the capabilities of traditional ensemble models.

Finally, extending the proposed framework beyond Kainji Lake National Park to other ecological zones in Nigeria and across West Africa would allow assessment of its transferability and scalability. Such efforts would support the development of regionally consistent wildfire early-warning systems, contributing to improved forest management, disaster risk reduction, and climate resilience planning at national and regional scales.

Compliance with ethical standards

The authors declare that the article did not violate any ethical standards within its jurisdiction

Disclosure of conflict of interest

The authors affirm that there is no conflict of interest to be disclosed.

References

[1] Adebowale, O. S., Oluleye, A., Olaiya, T., & Adedeji, O. (2023). Assessing the spatial and temporal distribution of fire regime in the derived savannah zone of Nigeria using space-based approach. World Scientific News, 183, 133–157. https://www.researchgate.net/publication/372885229

[2] Global Forest Watch. (2024). Nigeria — fires and tree cover loss dashboard. Retrieved from https://www.globalforestwatch.org/dashboards/country/NGA/?category=fires.

[3] Abulude, F. (2024). Forest fire in Nigeria. ResearchGate. https://doi.org/10.13140/RG.2.2.14660.85123

[4] Global Disaster Alert and Coordination System. (2025, January 5). Forest fire (17807 ha) in Nigeria 05 Jan 2025. https://www.gdacs.org/Wildfires/report.aspx?eventtype=WF&eventid=1023257

[5] Enoh, M. A., Okeke, U. C., & Narinua, N. Y. (2021). Identification and modelling of forest fire severity and risk zones in the Cross–Niger transition forest with remotely sensed satellite data. The Egyptian Journal of Remote Sensing and Space Science, 24(3, Part 2), 879–887. https://doi.org/10.1016/j.ejrs.2021.09.002

[6] Chen, Y., Morton, D. C., & Randerson, J. T. (2024). Remote sensing for wildfire monitoring: Insights into burned area, emissions, and fire dynamics. One Earth, 7(6), 1022–1028. https://doi.org/10.1016/j.oneear.2024.05.014

[7] Zhu, Z., Wang, S., & Woodcock, C. E. (2017). Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel-2 images. Remote Sensing of Environment, 159, 269–277. https://doi.org/10.1016/j.rse.2014.12.014

[8] Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., & Justice, C. O. (2018). The collection 6 MODIS burned area mapping algorithm and product. Remote Sensing of Environment, 217, 72–85. https://doi.org/10.1016/j.rse.2018.08.005

[9] Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J. L., Erb, A., Gao, F., Griffiths, P., Helder, D., Hostert, P., Huntington, J., Irons, J. R., Johnson, D. M., ... Zhu, Z. (2019). Current status of Landsat program, science, and applications. Remote Sensing of Environment, 225, 127–147. https://doi.org/10.1016/j.rse.2019.02.015

[10] Chen, X., Zhang, Y., Wang, S., Zhao, Z., Liu, C., & Wen, J. (2024). Comparative study of machine learning methods for mapping forest fire areas using Sentinel-1B and 2A imagery. Frontiers in Remote Sensing, 5, Article 1446641. https://doi.org/10.3389/frsen.2024.1446641

[11] Penot, V., Opitz, T., Pimont, F., & Merlin, O. (2025). Improved fire severity prediction using pre-fire remote sensing and meteorological time series: Application to the French Mediterranean area. Agricultural and Forest Meteorology, 371, 110588. https://doi.org/10.1016/j.agrformet.2025.110588

[12] Musa, A., Bello, A. A., & Shehu, A. (2023). Forest fire prediction based on time series networks and remote sensing images.7 MDPI.

[13] UNEP-WCMC & IUCN. (2023). Protected Planet: Kainji Lake National Park. https://www.protectedplanet.net

[14] FAO. (2022). Fire management – Global assessment 2021. Food and Agriculture Organization of the United Nations https://www.fao.org

[15] Beck, H. E., Zimmermann, N. E., McVicar, T. R., et al. (2018). Present and future Köppen–Geiger climate classification maps at 1-km resolution. Scientific Data, 5, 180214.

[16] Ibrahim, M., Balarabe, M., & Umar, A. (2022). Forest fire susceptibility prediction based on machine learning models with resampling algorithms on remote sensing data. OUCI.

[17] Abatzoglou, J. T., Williams, A. P., Boschetti, L., Zubkova, M., & Kolden, C. A. (2019). Global patterns of interannual climate–fire relationships. Global Change Biology, 24(11), 5164–5175. https://doi.org/10.1111/gcb.14405

[18] Archibald, S., Hempson, G. P., & Lehmann, C. E. R. (2019). A unified framework for plant life-history strategies shaped by fire and herbivory. New Phytologist, 224(4), 1490–1503.

[19] Andela, N., Morton, D. C., Giglio, L., et al. (2022). Global trends in burned area and fire emissions. Reviews of Geophysics, 60(1), e2020RG000726.

[20] Forkel, M., Dorigo, W., Lasslop, G., et al. (2019). Recent global and regional trends in burned area and their compensating environmental controls. Environmental Research Communications, 1(5), 051005.

[21] Olajesu, Sunday & Akinyemi, Abiodun & Funmilayo, Oni & Akeem, Lameed. (2019). Population Density, Diversity and Abundance of Antelope Species in Kainji Lake National Park, Nigeria. Open Journal of Ecology. 09. 107-116. 10.4236/oje.2019.94009.