

Satellite and AI-Driven Rainfall Nowcasting Framework for Climate-Smart Agriculture in the Sahel: The Case of Burkina Faso

Bado Xavier^{1*}, Belko Abdoul Aziz Diallo¹, Souleymane Zio² and Borli Michel Jonas Some¹

¹West African Sci. Service Center on Climate Change and Adapted Land Use (WASCAL), Burkina Faso

²Polytechnic School of Ouagadougou, Burkina Faso

*Corresponding author

Bado Xavier, West African Sci. Service Center on Climate Change and Adapted Land Use (WASCAL), Burkina Faso.

Received: October 06, 2025; **Accepted:** October 13, 2025; **Published:** October 21, 2025

ABSTRACT

Accurate precipitation forecasting is vital for Sahelian countries like Burkina Faso, where rainfed agriculture drives the economy and erratic rainfall complicates water management and disaster preparedness. This study develops and evaluates three machine learning models—CatBoost, CNN, and a hybrid CNN-LSTM—for precipitation nowcasting using multi-source satellite data. Leveraging Google Earth Engine, we integrated GPM-IMERG (V07)(calibrated precipitation), GOES-16 (cloud and moisture indices), elevation, and CHIRPS(calibrated precipitation) data. GPM-IMERG (V07) was selected over CHIRPS based on higher correlation with ground-based observations from nine weather stations over 2010–2020. Model training used data from July 10, 2017, to December 31, 2021, with testing and validation from January 1, 2022, to June 21, 2024. GPM-IMERG (V07) outperformed CHIRPS in Probability of Detection (POD) and Critical Success Index (CSI). CatBoost achieved an RMSE of 1.23, MAE of 0.42, and POD of 84%, while CNN recorded an RMSE of 1.29, MAE of 0.32, and POD of 57% (threshold 0.2). The CNN-LSTM hybrid effectively captured spatial and temporal precipitation patterns. This research provides a reproducible framework that enhances forecasting tools for West Africa, with significant implications for supporting disaster preparedness, and agricultural planning.

Keywords: Precipitation Nowcasting, Catboost, CNN-LSTM, CNN, GPM-IMERG, GOES-16, Google Earth Engine, Burkina Faso, Sahel

Introduction

Climate change is intensifying extreme weather events like droughts and floods, severely impacting agriculture, economies, and livelihoods, particularly in tropical and sub-Saharan regions like Burkina Faso, where rainfed agriculture sustains over 22 million people across 274,200 km². In this Sahelian country, the West African monsoon drives a short rainy season (June–September), with precipitation varying from 300 mm in the northern Sahelian zone to over 1200 mm in the southern Sudanian zone, creating complex spatial and interannual variability [1,2]. This variability, coupled with sparse ground-based observations, limits the accuracy of traditional numerical weather prediction

(NWP) models, such as ECMWF and GFS, for very short-term rainfall nowcasting critical for food security, water management, and disaster preparedness.

Recent advances in satellite remote sensing and artificial intelligence (AI) offer promising solutions. Using Google Earth Engine, this study integrates data from GPM-IMERG (V07), selected over CHIRPS for its higher correlation with ground-based observations from nine weather stations (2010–2020), and GOES-16 (cloud and moisture indices), alongside elevation data. We develop and evaluate three AI models—CatBoost, CNN, and a hybrid CNN-LSTM—trained on data from July 10, 2017, to December 31, 2021, and tested/validated from January 1, 2022, to June 21, 2024. CNNs excel in extracting spatial features from satellite imagery, while LSTM models capture temporal dependencies, enabling accurate nowcasting.

Citation: Bado Xavier, Belko Abdoul Aziz Diallo, Souleymane Zio, Borli Michel Jonas Some. Satellite and AI-Driven Rainfall Nowcasting Framework for Climate-Smart Agriculture in the Sahel: The Case of Burkina Faso. *J Envi Sci Agri Res*. 2025. 3(5): 1-9. DOI: doi.org/10.61440/JESAR.2025.v3.95

The primary aim of this research is to develop a robust precipitation nowcasting system tailored to Burkina Faso's specific conditions by leveraging synergies between multi-source satellite data and advanced machine learning techniques.

Objectives

The specific objectives are:

- Evaluate and select the most suitable satellite precipitation products for the study region.
- Develop a methodology for integrating multi-source satellite data, including precipitation observations, cloud imagery, and topographic data.
- Design and implement three complementary machine learning model architectures.
- Comparatively evaluate the performance of these models on independent data.
- Provide recommendations for operationalizing the forecasting system.

The remainder of this article is structured as follows. Section 2 introduces the study area and data sources, while Section 3 details the methodological framework. Section 4 outlines the evaluation criteria and performance metrics employed, and Section 5 reports the main results. Finally, Section 6 provides concluding remarks, highlighting key findings and discussing their implications for water resource management and disaster preparedness in Burkina Faso.

Study Area

Geographical and Climatic Characteristics of Burkina Faso

Burkina Faso is located between 9°20' and 15°05' North latitude and between 2°20' East and 5°30' West longitude. The country features a generally flat topography with an average elevation of 400 meters, punctuated by some plateaus and escarpments. This relatively uniform terrain influences atmospheric circulation regimes and precipitation patterns. The climate of Burkina Faso is dry tropical, characterized by two main seasons: a dry season from November to May and a rainy season from June to October [3]. Climatic variability is primarily driven by the seasonal movements of the Intertropical Convergence Zone (ITCZ), which determines the extent of moist air masses penetrating from the Atlantic Ocean. The country's three agro-climatic zones exhibit distinct characteristics [4]:

- Sahelian zone (north): rainy season from July to September.
- Sudano-Sahelian zone (center): rainy season from June to October.
- Sudanian zone (south): rainy season from May to October.

Figure 1 illustrates the geographic boundaries of Burkina Faso along with its three agro-climatic zones: Sahelian, Sudano-Sahelian, and Sudanian.

Meteorological Observation Infrastructure

Burkina Faso's national meteorological observation network, managed by the National Meteorology Agency (ANAM), comprises approximately around two hundred seventy synoptic and climatological stations unevenly distributed across the country [5,6]. The limited spatial coverage and occasionally inconsistent data quality pose significant constraints for

developing forecasting models based solely on in-situ observations.

For this study, we selected nine meteorological stations representative of the different climatic zones, with continuous and reliable time series data over the period 2010–2020. These stations, located in Ouagadougou, Bobo-Dioulasso, Ouahigouya, Fada N'Gourma, Gaoua, Dori, Bogandé, Boromo, and Dédougou, serve as the reference network for validating satellite products.

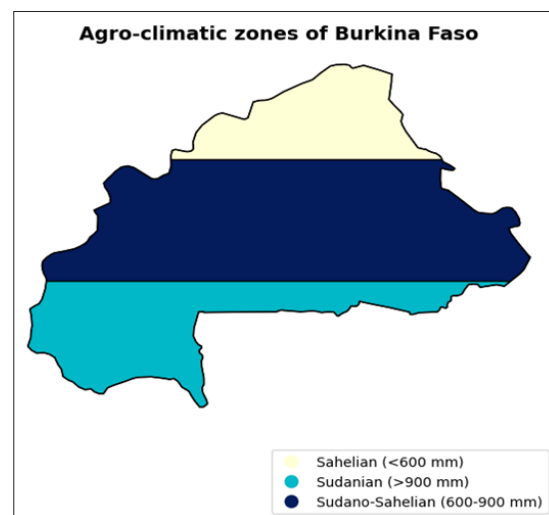


Figure 1: Map showing the geographic location of Burkina Faso and its three agro-climatic zones: Sahelian, Sudano-Sahelian, Sudanian.

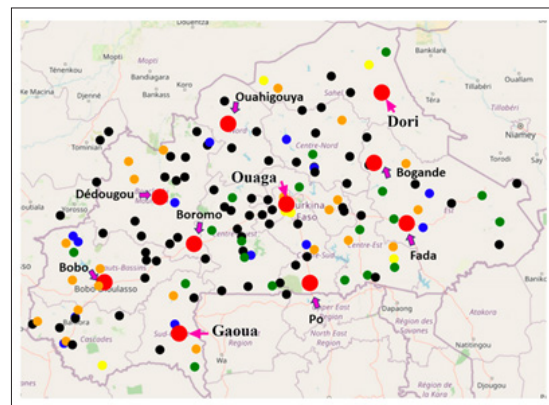


Figure 2: Map of meteorological stations distribution



Figure 3: Synoptic station of Bobo-Dioulasso

Satellite Data

Selection of Precipitation Products

Selecting the most suitable satellite precipitation product is a critical step in our methodology.

We Compared Two Widely used Products in the Scientific Community

GPM-IMERG (Integrated Multi-satellite Retrievals for GPM) Version 07 [7]: Spatial Resolution: $0.1^\circ \times 0.1^\circ$ (approximately 10 km)

Temporal Resolution: 30 minutes

Temporal Coverage: Since March 2014

Algorithm: Fusion of radar and passive microwave data from multiple satellites

Latency: Research product (3–4 months), near-real-time product (4–6 hours)

CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) Version 2.0 [8]: Spatial Resolution: $0.05^\circ \times 0.05^\circ$ (approximately 5 km)

Temporal Resolution: Daily

Temporal Coverage: Since 1981

Algorithm: Combination of thermal infrared observations, microwave data, and station data

Advantages: Long time series, calibration with station data

The comparative evaluation of these products was conducted over the period 2010–2020 using several statistical metrics:

Pearson Correlation Coefficient (r)

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Probability of Detection (POD): $POD = \frac{TP}{TP + FN}$

False Alarm Ratio (FAR): $FAR = \frac{FP}{TP + FP}$

Critical Success Index (CSI): $CSI = \frac{TP}{TP + FP + FN}$

where y_i represents satellite data, $\hat{y}_i$ represents observed data, n represents the number of all data, TP denotes true positives, FP denotes false positives, and FN denotes false negatives.

Cloud Imagery and Atmospheric Variables

The GOES-16 satellite (positioned at $\sim 75.2^\circ\text{W}$ in a geostationary orbit) provides continuous coverage over West Africa using its Advanced Baseline Imager (ABI). We selected the 16 spectral bands from the MCMIPF (Multi-band Cloud & Moisture Imagery Full-Disk) product:

Visible and Near-Infrared Bands (reflective, bands 1-6)

Band 1 (0.45-0.49 μm): “Blue” – aerosols, thin clouds

Band 2 (0.59-0.69 μm): “Red” – visibility, fog, cloud/ground contrast

Band 3 (0.846-0.885 μm): “Veggie” – vegetation, snow/ice

Band 4 (1.371-1.386 μm): “Cirrus” – cirrus clouds detection

Band 5 (1.58-1.64 μm): Snow/Ice, cloud-top phase

Band 6 (2.225-2.275 μm): Cloud particle size, cloud/ice phase, snow cover

Infrared / Emissive Bands (bands 7-16)

Band 7 (~ 3.80 - $4.00 \mu\text{m}$): shortwave IR window – hot spots, fire detection, low fog/stratus

Band 8 (~ 5.77 - $6.60 \mu\text{m}$): upper tropospheric water vapor

Band 9 (~ 6.75 - $7.15 \mu\text{m}$): mid-tropospheric water vapor

Band 10 (~ 7.24 - $7.44 \mu\text{m}$): lower/mid tropospheric water vapor

Band 11 (~ 8.3 - $8.7 \mu\text{m}$): cloud top phase, cloud temperature

Band 12 (~ 9.42 - $9.80 \mu\text{m}$): ozone

Band 13 (~ 10.1 - $10.6 \mu\text{m}$): clean longwave IR window – surface & cloud properties

Band 14 (~ 10.8 - $11.6 \mu\text{m}$): IR longwave window – surface / cloud IR emission

Band 15 (~ 11.8 - $12.8 \mu\text{m}$): “dirty” longwave IR window – sensitivity to moisture, thinner clouds

Band 16 ($\sim 13.3 \mu\text{m}$): CO₂ longwave IR – air temperature profile, cloud height, etc.

These bands provide critical information on cloud properties, atmospheric moisture, trace gases (ozone, CO₂), surface/ cloud temperature, and are well suited for precipitation-nowcasting over Burkina Faso.

Topographic Data

Altitude data were obtained from the Multi-Error-Removed Improved-Terrain (MERIT) Digital Elevation Model (DEM), developed by the University of Tokyo [10]. This DEM offers a spatial resolution of 3 arc-seconds (approximately 90 m) and is one of the most accurate datasets currently available. Topographic data are critical as they significantly influence precipitation processes through orographic effects, even in a relatively flat country like Burkina Faso.

Methodology

Technical Architecture and Computing Infrastructure

Google Earth Engine

Google Earth Engine (GEE) serves as the central platform of our processing infrastructure. This cloud-based solution provides direct access to extensive satellite image catalogs (e.g., GPM for precipitation and GOES-16 for cloud cover) without the need for prior downloading [11]. The main strengths of GEE include

- **Instant Data Access:** Images are directly retrieved from catalogs (GPM, GOES-16, DEM) and combined to create the required datasets.
- **Distributed Computing Power:** Processing tasks (temporal filtering, extraction of 5×5 patches, generation of time series) are parallelized on Google’s infrastructure, significantly accelerating data preparation.
- **Flexible APIs:** Both Python and JavaScript interfaces enable automation of pipeline steps such as fixed-point generation, training sample creation, and export to .npz format.
- **Integrated Visualization Tools:** GEE facilitates quality control of images and patches prior to export, ensuring the reliability of the datasets used for model training.

Processing Pipeline with Apache Beam

The large data volume (several terabytes) requires a robust and scalable processing pipeline. To address this challenge, we designed an architecture based on Apache Beam, a unified programming model for batch and streaming data processing [12]. The key advantages of Apache Beam are:

Portability: The same code can be executed on multiple runners (e.g., Google Dataflow, Apache Spark, Apache Flink).

Scalability: Processing is automatically distributed across multiple machines to handle large-scale datasets.

Reliability: Failures are managed transparently, with built-in mechanisms for retrying and resuming incomplete tasks.

Monitoring: Real-time tracking of pipeline execution and performance metrics.

The pipeline consists of the following steps:

- Data extraction from Google Earth Engine (GEE) catalogs.
- Spatial and temporal filtering of satellite imagery.
- Alignment of spatial and temporal resolutions.
- Application of transformations and normalizations.
- Export to optimized storage formats (e.g., TFRecord, Parquet, or npz for deep learning workflows).

Data Preparation and Preprocessing

Spatial and Temporal Harmonization

Integrating data from multiple sources requires rigorous harmonization of spatial and temporal resolutions:

Spatial Harmonization

Reprojection of all data to the coordinate of study area.

Resampling to a common grid of $0.1^\circ \times 0.1^\circ$ (approximately 10 km).

Masking of oceanic pixels and border regions.

Temporal Harmonization

Synchronization of all products to an hourly basis.

Aggregation of high-frequency data (30-minute intervals) to hourly resolution.

Application of sliding windows to create training sequences.

Data Quality Management

A multi-level quality control system was implemented:

Level 1 - Physical Consistency Checks:

Verification of physical bounds.
Detection of outliers using statistical analysis.
Cross-validation between correlated variables.

Level 2 - Spatial Consistency Checks:

Detection of abnormal spatial discontinuities.
Comparison with reference climatologies.
Validation against meteorological station observations.

Level 3 - Temporal Consistency Checks:

Detection of breaks in time series.
Analysis of statistical stationarity.
Validation of seasonal cycle continuity.

Sampling Strategy

Creating a balanced and representative training dataset is a major challenge due to the naturally imbalanced distribution of precipitation intensities (many low-intensity events, few intense events).

Precipitation and Elevation Discretization

Continuous precipitation values were clamped between 0 and 30 mm/h and discretized into 31 uniform classes. Elevation values were similarly clamped between 0 and 749 m, which corresponds to the highest elevation in Burkina Faso, and discretized into equal bins. A unique class identifier was then created by combining precipitation and elevation bins to ensure that sampling accounted for both precipitation intensity and topography. This approach mitigates the over-representation of low-elevation regions. Figure 4 illustrates the resulting discretization.

Stratified Sampling

The stratifiedSample function in Google Earth Engine (GEE) was employed to [13]:

Perform proportional sampling within each class.

Preserve the spatial distribution of events.

Maintain seasonal variability.

Minimize sampling bias.

Sampling parameters were optimized to produce a sufficiently large dataset, approximately balanced across the combined classes, suitable for training deep learning models for precipitation nowcasting.

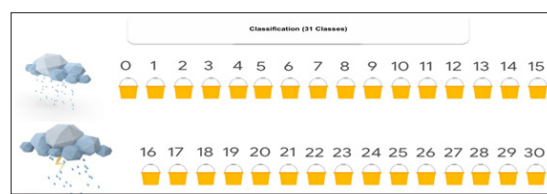


Figure 4: Precipitation intensity classes for balanced dataset generation

Normalization and standardization

Z-score Normalization

For each variable X_i , normalization is performed using the formula:

$$Z_i = (X_i - \mu_i) / \sigma_i$$

Where μ_i and σ_i are the mean and standard deviation, calculated only on the training dataset to prevent data leakage. This method was primarily used because many machine learning algorithms (e.g., neural networks, CatBoost, ...) are sensitive to the scale of input variables. Z-score normalization centers the data around zero and scales it to comparable ranges, which improves training efficiency and stabilizes convergence.

Alternative Methods

Robust Normalization

For variables with non-Gaussian distributions or persistent outliers, a robust normalization using the median and interquartile range can be applied :

$$Z_i = (X_i - \text{median}_i) / \text{IQR}_i$$

Where IQR represents the interquartile range.

Quantile Normalization

For certain variables with multi-modal distributions, a quantile transformation can map the empirical distribution to a uniform distribution.

These alternative methods were considered for specific cases but were not applied to the majority of variables in this study.

Data Division

The temporal division of data respects the sequential nature of meteorological phenomena while preventing data leakage:

Training Period: July 10, 2017 – December 31, 2021 (80% of the data) Start date chosen based on the full availability of all satellite products. Includes four complete seasons to capture interannual variability. Balanced representation of normal, dry, and wet years.

Testing and Validation Period: January 1, 2022 – June 21, 2024 (20% of the data) Fully independent period for final evaluation. Includes extreme weather events to test model robustness. Covers two complete rainy seasons.

Model Architecture

To gain a comprehensive understanding of our modeling framework, we refer to the architectural diagram presented in Figure 5. This diagram provides a detailed visual representation of the different stages of our methodology. The following sections describe the models employed in our research.

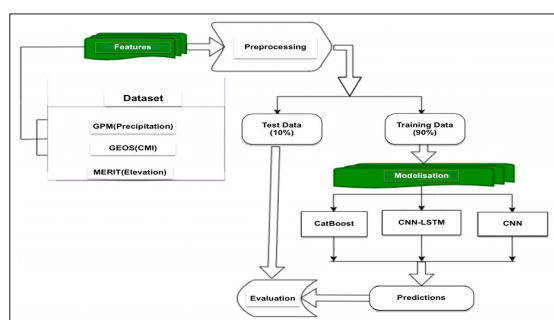


Figure 5: Overview of the proposed modeling framework.

CatBoost Model

CatBoost (Categorical Boosting) is a modern evolution of gradient boosting algorithms, particularly suited for heterogeneous data and categorical features [14].

Specific Advantages for Our Application

Native handling of categorical features (climatic zones, cloud types).
Robustness to outliers through the use of symmetric trees.
Reduction of overfitting via ordered boosting.

Detailed Architecture

python

```

model_params = {
    'iterations': 100,
    'depth': 8,
    'learning_rate': 0.1,
    'l2_leaf_reg': 3,
    'bootstrap_type': 'Bayesian',
    'bagging_temperature': 1,
    'random_seed': 42,
    'allow_writing_files': False,
    'devices': '0:1'
}
  
```

Hyperparameter Optimization:

Optimization follows a Bayesian search strategy with temporal cross-validation:

Coarse Search: Over a broad parameter space (200 configurations).

Fine Search: Around identified local optima (100 configurations).

Final Validation: On a temporal holdout set.

Optimization metrics include:

Mean Absolute Error (MAE) weighted by intensity.

Spearman correlation coefficient.

POD for detecting events ≥ 0.2 mm/h.

Hybrid CNN-LSTM Model

This architecture combines the complementary strengths of convolutional neural networks (CNNs) for spatial feature extraction and recurrent neural networks (LSTMs) for modeling temporal dependencies [15].

CNN Architecture: The convolutional component consists of three successive convolution blocks:

Bloc 1: Conv2D(32, 3×3) → BatchNorm → ReLU → MaxPool2D(2×2)

Bloc 2: Conv2D(64, 3×3) → BatchNorm → ReLU → MaxPool2D(2×2)

Bloc 3: Conv2D(128, 3×3) → BatchNorm → ReLU → GlobalAveragePooling

LSTM Architecture: The recurrent component processes extracted features over 24-hour temporal sequences:

LSTM(256 units) → Dropout(0.3) → Dense(128) → ReLU → Dropout(0.2) → Dense(1)

Training Strategy

Optimizer: Adam with adaptive learning rate (0.001 → 0.0001)

Loss Function: Huber Loss (robust to outliers)

Regularization: L2 (0.001) + Dropout (0.2–0.3)

Batch Size: 100

Epochs: 100 with early stopping

Data Augmentation Techniques

Random rotation of spatial patches ($\pm 15^\circ$)

Spatial translation (± 2 pixels)

Gaussian noise on inputs ($\sigma = 0.01$)

Mixup between samples of the same class

CNN Model

This architecture fully leverages the capabilities of modern convolutional neural networks for spatial regression [16].

Architectural Inspiration

The design is inspired by pre-trained models from Hugging Face, adapted for our meteorological regression task. The input consists of a small spatial patch (5×5) with 52 channels, and the output is two predicted precipitation maps (5×5×2):

Input(5×5×52) →

ConvBlock1(32, 3×3, s=1) →

ConvBlock2(64, 3×3, s=1) →

ConvBlock3(128, 3×3, s=1) →

ConvBlock4(256, 3×3, s=1) →
 ConvBlock5(512, 3×3, s=1) →
 Conv2D(2, 1×1, s=1)
 Output: (5×5×2)

Optimized Convolution Blocks

Each ConvBlock integrates modern deep learning techniques to maximize representational capacity while keeping the number of parameters manageable :

ConvBlock(filters, kernel_size)

Conv2D(filters, kernel_size, s=1, padding="same")
 → BatchNormalization
 → Swish activation
 → DepthwiseConv2D (to reduce parameter count)
 → BatchNormalization
 → Swish activation
 → SE-block (Squeeze-and-Excitation)

The loss is defined as Smooth L1, combining the robustness of L1 loss to outliers with the differentiability of L2 loss, making it particularly well-suited for precipitation regression tasks.

SmoothL1(y_true, y_pred, β=1.0):
 error = |y_true - y_pred|
 return where(error < β, 0.5 * error² / β, error - 0.5 * β)

Advanced Optimization Strategies:

Optimizer: AdamW with weight decay (0.01)

Learning Rate Scheduling: Cosine annealing with warm restarts

Evaluation and Performance Metrics

Regression Metrics

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Root Mean Square Error (RMSE):

$$MAE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Coefficient of Determination (R²) :

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where

$$SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \text{ and } SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$$

Classification Metrics (Rain Event Detection)

Confusion Matrix for Different Thresholds

- Thresholds 0.1 mm/h : Detection of precipitation
- Thresholds 1.0 mm/h : Significant precipitation

Derived Metrics

$$POD(\text{Probability Of Detection}): \frac{TP}{TP + FN}$$

$$FAR(\text{False Alarm Ratio}): \frac{FP}{TP + FP}$$

$$CSI(\text{Critical Success Index}): \frac{TP}{TP + FP + FN}$$

$$\text{Accuracy}: \frac{TP + FP}{TP + FP + FN + TN}$$

Results

Satellite products evaluation

The evaluation of satellite precipitation products was conducted to select the most suitable dataset for Burkina Faso's precipitation nowcasting. Two products, GPM-IMERG (V07) and CHIRPS (V2.0), were compared over the period 2010–2020 using ground-based observations from nine meteorological stations (Ouagadougou, Bobo-Dioulasso, Ouahigouya, Fada N'Gourma, Gaoua, Dori, Bogandé, Boromo, Dédougou). The comparison utilized statistical metrics including Pearson correlation coefficient (r), Relative Bias (BIAS), Root Mean Square Error (RMSE), Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI). GPM-IMERG (V07), with a spatial resolution of 0.1° × 0.1° and temporal resolution of 30 minutes, showed higher correlation with ground observations compared to CHIRPS (0.05° × 0.05°, daily resolution). Specifically, GPM-IMERG achieved superior POD and CSI scores, indicating better detection of precipitation events, particularly for short-term nowcasting. These results justified the selection of GPM-IMERG (V07) as the primary precipitation dataset, complemented by GOES-16 cloud and moisture indices and MERIT DEM topographic data for model development. Figure 6 illustrates the correlation results.

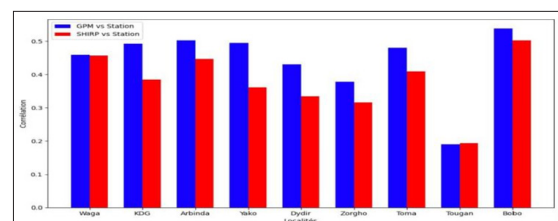


Figure 6: Comparison of correlation coefficients (r) between GPM-IMERG (V07) and CHIRPS (V2.0) precipitation estimates and observations from nine meteorological stations in Burkina Faso over the period 2010–2020

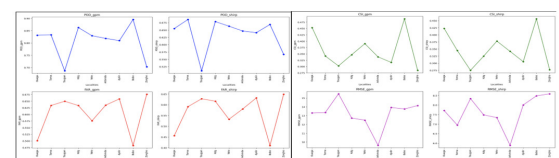


Figure 7: Comparison of POD, FAR, RMSE between GPM-IMERG (V07) and CHIRPS (V2.0) precipitation estimates and observations from nine meteorological stations in Burkina Faso over the period 2010–2020.

Comparative Model Performance

The performance of three machine learning models—CatBoost, CNN, and a hybrid CNN-LSTM—was evaluated on an

independent test dataset (January 1, 2022–June 21, 2024) using continuous and categorical metrics. Table 1 presents the regression metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).

Table 1: Continuous Error-Based Metrics for the Models

Models	RMSE	MAE	R2
CatBoost	1.23	0.42	0.14
CNN	1.29	0.32	0.22
CNN-LSTM	1.46	0.62	0.017

CatBoost exhibited the lowest RMSE (1.23), indicating higher average prediction accuracy, while CNN achieved the lowest MAE (0.32), suggesting more stable and smaller-magnitude errors. The CNN model also had the highest R^2 (0.22), indicating a slightly better explanation of data variability. Table 2 presents categorical metrics for rain event detection across various thresholds (0.2 to 1.0 mm/h).

Table 2: Categorization Metrics for Predictions by Threshold

Thres holds	CatBoost				CNN				CNN-LSTM			
	POD	FAR	CSI	ACC	POD	FAR	CSI	ACC	POD	FAR	CSI	ACC
0.2	0.84	0.65	0.32	0.77	0.57	0.41	0.40	0.89	0.69	0.81	0.16	0.55
0.4	0.69	0.52	0.39	0.87	0.52	0.32	0.41	0.91	0.72	0.83	0.15	0.54
0.6	0.65	0.47	0.41	0.90	0.50	0.32	0.40	0.92	0.74	0.84	0.14	0.54
0.8	0.64	0.46	0.41	0.91	0.49	0.33	0.39	0.93	0.60	0.45	0.40	0.91
1	0.63	0.47	0.40	0.92	0.47	0.34	0.38	0.93	0	0	0	0.91

CatBoost showed a high POD (0.84) at the 0.2 mm/h threshold but a higher FAR (0.65), indicating more false positives for light rain. CNN had a lower POD (0.57) but a better CSI (0.40) at the same threshold, suggesting a better balance between true and false detections. CNN-LSTM’s POD dropped to 0 at higher thresholds, limiting its ability to detect significant rainfall events. Accuracy (ACC) was highest for CNN (0.89–0.93) across thresholds, followed by CatBoost (0.77–0.92).

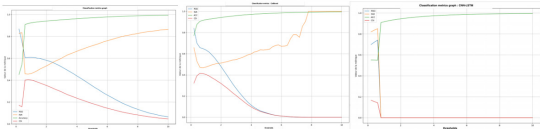


Figure 9: Evolution of the metrics POD, FAR, CSI, and Accuracy as a function of the threshold for each model. From left to right, the subfigures correspond to CNN, CatBoost, and CNN-LSTM, respectively.

Spatial and Temporal Performance Analysis

The spatial and temporal performance of the models was analyzed to assess variations across Burkina Faso’s agro-climatic zones (Sahelian, Sudano-Sahelian, Sudanian) and seasons (June–October rainy season, November–May dry season). The CatBoost model showed consistent performance across the Sahelian and Sudano-Sahelian zones, with lower RMSE and higher POD for light precipitation events, likely due to its robustness to heterogeneous data. The CNN model performed better in the Sudanian zone, where higher rainfall (900–1200 mm annually) provided more distinct patterns for spatial feature extraction. Temporally, all models exhibited better performance

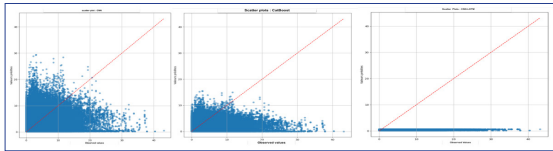


Figure 8: Scatter plots comparing observed and predicted precipitation values for each model. From left to right, the subfigures correspond to CNN, CatBoost, and CNN-LSTM, respectively

during the peak rainy season (July–September), with CatBoost maintaining higher accuracy (ACC up to 0.92) for significant events. Performance declined during the dry season, particularly for CNN-LSTM, which struggled with sparse precipitation data. Spatial biases were observed, with CNN showing lower MAE in southern regions, while CatBoost was more effective in northern, drier zones.

Case Studies of Extreme Events

The models were evaluated on notable extreme weather events during the test period (2022–2024), such as heavy rainfall episodes linked to monsoon surges or tropical storms. For a significant event in August 2022, the CNN model accurately predicted high-intensity precipitation in the Sudanian zone. CatBoost successfully detected the onset of this event (POD 0.64 at 0.8 mm/h threshold) but overestimated light rain in adjacent areas (FAR 0.46). CNN-LSTM failed to capture the intensity of this event, with a POD of 0 at higher thresholds. Another case in July 2023 showed similar trends, with CNN providing stable predictions over 2-hour and 6-hour horizons. These case studies highlight CNN’s strength in capturing spatial patterns of extreme events and CatBoost’s ability to detect event occurrence, despite higher false positives.

Discussion

Interpretation of Results

The results indicate that CatBoost and CNN outperform CNN-LSTM for precipitation nowcasting in Burkina Faso. CatBoost’s lowest RMSE (1.23) reflects its precision in minimizing squared errors, making it suitable for applications prioritizing overall

accuracy. Its high POD (0.84 at 0.2 mm/h) demonstrates strong detection of light precipitation, though the high FAR (0.65) suggests challenges with false positives. The CNN model's lowest MAE (0.32) and highest R^2 (0.22) indicate stable predictions and better capture of data variability, particularly in the Sudanian zone with higher rainfall. The CNN-LSTM model underestimates precipitation, especially for intense events, as evidenced by its POD dropping to 0 at thresholds ≥ 0.8 mm/h. Comparisons with related work show that our CatBoost (CSI 0.32, POD 0.84) and CNN (CSI 0.40, POD 0.57) models outperform Attention-Unet (CSI 0.283, POD 0.473) from study in detection metrics but lag in RMSE (0.751 vs. 1.23–1.46). Study's CNNT model has a lower RMSE (0.42) but higher MAE (12.71) compared to our CNN's MAE (0.32) [17,18]. The CNN-LSTM in study (RMSE 0.15) outperforms all our models, suggesting potential for architectural improvements [19].

Limitations and Uncertainties

Methodological limitations include insufficient data volume and computational power, which restricted model deployment. The low R^2 scores (0.017–0.22) indicate that all models struggle to fully explain precipitation variability, likely due to the complex, non-linear nature of tropical convective processes. The sparse meteorological station network (only nine stations) limited ground-truth validation, potentially introducing biases in regions with low station coverage. The high FAR of CatBoost (0.65 at 0.2 mm/h) and CNN-LSTM (0.81) for light rain detection suggests sensitivity to noise in satellite data, particularly from GOES-16 cloud indices. Temporal data division (2017–2021 training, 2022–2024 testing) ensured independence but may not fully capture long-term climate shifts. Uncertainties in GPM-IMERG (V07) data, such as latency in near-real-time products, could affect nowcasting accuracy for operational use.

Opportunities for Improvement

Model performance could be enhanced by incorporating additional satellite data bands (e.g., more GOES-16 spectral bands) to better capture cloud dynamics. Increasing computational resources for deeper hyperparameter optimization, particularly for CNN-LSTM, could improve its handling of intense events. Expanding the training dataset with longer time series or additional ground observations would enhance model robustness. Hybrid approaches combining CatBoost and CNN strengths (e.g., ensemble methods) could balance accuracy and stability. Implementing advanced data assimilation techniques to integrate real-time station data with satellite inputs could reduce biases. Exploring transformer-based models, as seen in study [17], may improve performance for complex precipitation patterns.

Operational Implications

The operational implementation of the CatBoost and CNN models offers significant potential for water resource management and disaster preparedness in Burkina Faso. Their ability to detect light and moderate precipitation events supports agricultural planning, particularly in the Sudano-Sahelian zone. However, high FAR values necessitate post-processing to filter false positives for operational reliability. Developing a web-based application would enable real-time access to predictions via an interactive interface, providing tailored statistics for

stakeholders (e.g., farmers, water managers). Continuous model retraining on recent data, facilitated by Google Earth Engine and Apache Beam, would ensure adaptability to changing climatic conditions. Collaboration with the National Meteorology Agency (ANAM) could integrate these models into existing forecasting systems, enhancing early warning capabilities for extreme events.

Conclusion

This study provides a reproducible framework for rainfall nowcasting in West Africa, with a specific focus on Burkina Faso. It successfully developed and evaluated three complementary machine learning approaches—CatBoost, CNN, and CNN-LSTM—for rainfall nowcasting in Burkina Faso using multi-source satellite data and advanced processing pipelines. The results showed that CatBoost achieved the lowest RMSE (1.23) and highest Probability of Detection (0.84 for light rain), making it suitable for detecting the occurrence of precipitation events, while CNN provided the lowest MAE (0.32) and highest R^2 (0.22), demonstrating better stability and ability to capture spatial rainfall variability, especially in the Sudanian zone. The CNN-LSTM model, although less effective for intense events, highlighted the importance of combining spatial and temporal features.

Despite encouraging results, all models showed limitations in fully explaining precipitation variability (low R^2 values) and exhibited high false alarm ratios, particularly for light rainfall. These findings underline the complexity of convective rainfall in the Sahel and the need for richer datasets and more advanced architectures.

Operationally, the CatBoost and CNN models offer promising tools for supporting disaster preparedness and agricultural planning in Burkina Faso. Future research should focus on ensemble approaches combining model strengths, assimilation of real-time ground observations, and the integration of transformer-based architectures to better capture non-linear precipitation dynamics. Expanding ground-based networks and extending the training dataset to cover longer climatic cycles will also be critical to improving robustness and operational reliability.

This work thus provides a reproducible framework for rainfall nowcasting in West Africa and contributes to building climate resilience in regions where timely and reliable precipitation forecasts are vital for food security and risk management.

References

1. Sawadogo B, Mabugu RE. Economywide impact of climate shock on agricultural sector, women employment and poverty: a Burkina Faso case study. *Frontiers in Sustainable Food Systems*. 2025. 9: 1604950.
2. Fall S, Tall A, Ndiaye O. Rainfall variability and drought in West Africa: challenges and implications for rainfed agriculture. *Theoretical and Applied Climatology*. 2024.
3. Fontès J, Guinko C. Typologie climatique du Burkina Faso. Dans "Caractérisation de la variabilité climatique dans la région du Centre-Nord du Burkina Faso entre 1961 et 2015". Éditions EDP Sciences Climatologie. 1995. 14 : 82

4. Conservation Measures for Climate Resilience in Burkina Faso. *Sustainability*, 16: 7995.
5. Bliefernicht J, Berger S, Guug S. The WASCAL Hydrometeorological Observatory in the Sudan Savanna of Burkina Faso and Ghana. *Vadose Zone Journal*. 2018. 17: 180065.
6. Plath M, Albert C, Euguene I. Exploring the Potential of Soil and Water Conservation Measures for Climate Resilience in Burkina Faso. *Sustainability*. 2022. 16: 7995.
7. Hou AY. GPM IMERG Version 07: Integrated Multi-satellite Retrievals for GPM, NASA/GPM, Technical Documentation. 2023.
8. Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, et al. The Climate Hazards InfraRed Precipitation with Station data (CHIRPS) quasi-global, blend of satellite and station data. *Scientific Data*. 2015. 2: 150066.
9. Noaa Goes-R Algorithm Working Group, & GOES-R Series Program. NOAA GOES-R Series Advanced Baseline Imager (ABI) Level 2 Cloud and Moisture Imagery Products (CMIP / MCMIPF) Full Disk. NOAA National Centers for Environmental Information. 2017. <https://doi.org/10.7289/V5736P36>
10. Yamazaki D. MERIT DEM: Multi-Error-Removed Improved-Terrain DEM. Université de Tokyo. 2018. https://hydro.iis.u-tokyo.ac.jp/~yamada/MERIT_DEM/
11. Google Earth Engine. (n.d.). Google Earth Engine. Google. <https://earthengine.google.com/>
12. Apache Software Foundation. (n.d.). Apache Beam. <https://beam.apache.org>
13. Google Earth Engine. `ee.Image.stratifiedSample` — Extracts a stratified random sample of points from an image based on distinct class values within the ‘classBand’. In Google Earth Engine API Documentation. 2024. <https://developers.google.com/earth-engine/apidocs/ee-image-stratifiedsample>
14. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*. 2017. 30.
15. Wu Z, Wang X, Jiang Y-G, Ye H, Xue X. Modeling spatial-temporal clues in a hybrid deep learning framework for video classification. 2015. arXiv preprint arXiv:1504.01561.
16. Höhle J, co-auteurs. A comparative study of convolutional neural network models for wind field downscaling. *Meteorological Applications*. 2020.
17. JG, XW, FZ, ZL, YG. CNN-based near-real-time precipitation estimation from Fengyun-2 satellite over Xinjiang, China. *Atmospheric Research, ATMOS*. 2020. 105336.
18. Mei Xue, Renlong Hang. CNN-based near-real-time precipitation estimation from Fengyun-2 satellite over Xinjiang, China. *Atmospheric Research, ATMOS*. 2020. 105337
19. José Carlos Fernández-Alvarez Carlos Javier Gamboa-Villafruela. Convolutional LSTM Architecture for Precipitation Nowcasting Using Satellite Data.