Research Square

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

Reproducible Hybrid Ensemble Learning for Drought Forecasting

Moses Chilanga

moseschilanga22@gmail.com

Copperbelt University

Josephat Kalezhi

Copperbelt University

Nchimunya Chaamwe

Copperbelt University

Research Article

Keywords: Drought forecasting, Machine learning models, Hybrid ensemble methods, Feature engineering, Climate resilience in Zambia, Early warning systems

Posted Date: September 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10781249/v1

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

Additional Declarations: No competing interests reported.

Reproducible Hybrid Ensemble Learning for Drought Forecasting

Moses Chilanga[1*], Josephat Kalezhi[2], and Nchimunya Chaamwe[3].

[1] Information Technology and communication, Copperbelt University, Kitwe, Zambia
[2] Information Technology and communication, Copperbelt University, Kitwe, Zambia
[3] Information Technology and communication, Copperbelt University, Kitwe, Zambia

Corresponding author: Moses Chilanga (e-mail: moseschilanga22*@gmail.com*).

Abstract: Drought forecasting in Zambia is critical for agricultural resilience and climate adaptation. This study introduces a reproducible hybrid ensemble learning framework that integrates classical machine learning algorithms (Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine, K-Nearest Neighbors, Decision Tree) with deep learning architectures (Artificial Neural Network, Long Short-Term Memory, Convolutional Neural Network). Using Zambia's climate dataset (2015–2025; 12,000 daily records of rainfall, temperature, soil moisture, and derived indices), the framework applies domain-driven feature engineering including lag variables, rolling averages, and interaction terms to embed hydrological memory and cumulative drought dynamics. A five-fold temporal cross-validation demonstrated that Gradient Boosting and the Hybrid Ensemble achieved comparable performance (95% accuracy, balanced precision and recall, $p > 0.05$). Importantly, the contribution lies not in surpassing Gradient Boosting but in providing an adaptive, reproducible, and domain-informed ensemble pipeline. The adaptive weighting mechanism combines Random Forest feature importance with CHIRPS rainfall validation, while reproducibility is ensured through Google Colaboratory, GitHub version control, controlled random seeds, and CRISP-DM documentation. Confusion matrix analysis highlighted challenges in distinguishing moderate from severe drought classes, underscoring the value of engineered features. By embedding satellite rainfall (CHIRPS) and soil moisture proxies (MODIS), the framework demonstrates methodological transparency, scalability, and operational relevance, positioning drought forecasting as a case study in reproducible ensemble learning and computing innovation.

Keywords: Drought forecasting; Machine learning models; Hybrid ensemble methods; Feature engineering; Climate resilience in Zambia; Early warning systems.

1 INTRODUCTION

Zambia's vulnerability to recurrent droughts has intensified in recent decades, severely affecting crop yields, water resources, and rural livelihoods. Droughts in Southern Africa are strongly linked to El Niño events and long-term climate variability, often resulting in food insecurity and economic losses (Wang et al. 2022). Accurate and timely drought forecasting is therefore essential for strengthening national early warning systems and guiding climate adaptation strategies. Traditional statistical approaches, such as regression and autoregressive models, often fail to capture nonlinear climate interactions and multiscale dependencies, motivating the adoption of machine learning (ML) techniques (Lee et al. 2025).

Recent advances in ML have demonstrated strong potential for environmental forecasting. Ensemble methods such as Random Forests and Gradient Boosting have proven effective in capturing complex hydrological dynamics and reducing over fitting (Tyagi et al. 2022). Deep learning approaches, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have been applied to drought prediction in Africa and Asia, showing promise but often constrained by limited dataset size and temporal depth (Tyagi et al. 2022). However, systematic reviews confirm that while ML applications in drought forecasting are expanding rapidly, reproducibility and methodological novelty remain weak points (Mokhtar et al. 2021). Pipelines are often poorly documented, ensemble methods are rarely integrated with reproducibility practices, and few studies demonstrate methodological scalability in data-constrained environments (Wang et al. 2022).

This paper addresses that gap by introducing a reproducible hybrid ensemble pipeline for drought forecasting in Zambia. Unlike prior studies that focus on single models or binary drought classification, we integrate classical algorithms, deep learning architectures, and a novel hybrid ensemble framework within a reproducibility-driven workflow. The pipeline embeds transparent documentation, controlled randomness, and version-controlled code into ensemble learning, ensuring that results are not only accurate but also replicable and auditable (Kumar et al. 2026). Domain-based feature engineering introduces lag variables, rolling averages, and interaction terms to embed hydrological memory and cumulative drought dynamics into the dataset. By applying rigorous validation (five-fold cross-validation, confusion matrix analysis, and statistical testing), this study demonstrates how reproducibility and ensemble innovation together advance computing methodology while providing a foundation for future climate resilience research.

2 MATERIALS AND METHODS

2.1 DATASET

The dataset employed in this study integrates ground-based meteorological observations with satellite-derived climate products, ensuring both temporal continuity and spatial completeness. Primary data were obtained from the Zambia Meteorological Department (ZMD), which provided daily records of rainfall, temperature, soil moisture, and drought indices. To address spatial gaps in station coverage, the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset was incorporated, offering precipitation estimates at 0.05° resolution across Zambia and thereby reducing biases inherent in sparse ground networks (Hukkeri et al. 2023). Soil moisture proxies were supplemented with MODIS satellite products, which capture vegetation and hydrological stress responses and have been widely applied in drought monitoring (Mokhtari and Akhoondzadeh 2020).

The study period spans 2015–2025, covering a full decade of climate variability. This timeframe includes both El Niño and La Niña events, ensuring representation of diverse drought conditions and extremes that have historically affected Southern Africa (Deo and Şahin 2015). In total, approximately 12,000 daily records were compiled, with 15 predictor features. These features included raw meteorological variables (rainfall, temperature, soil moisture), engineered lag and rolling features to capture cumulative drought dynamics, and derived indices such as rainfall z-scores and SPEI proxies to embed established climatological measures directly into the dataset (Nahidi et al. 2026).

The target variable was defined as drought severity, categorized into three classes: No drought, Moderate drought, and severe drought. Classification thresholds were based exclusively on the Standardized Precipitation Index (SPI), a widely recognized measure of hydrological drought severity across multiple timescales (Zhang et al. 2020). The forecasting models in this study predict SPI-based drought severity classes only. The Normalized Drought Index (NDI) was introduced solely as a supporting diagnostic tool to validate predictor associations and feature relevance; it was never used as the dependent variable in forecasting. The class distribution was moderately imbalanced, with 60% of records classified as No drought, 25% as Moderate drought, and 15% as severe drought. This imbalance reflects Zambia's climatology, where severe droughts are less frequent but highly impactful. To address this, weighted metrics such as precision, recall, and F1 score were reported alongside accuracy, ensuring fairness in model comparison (Sundararajan et al. 2021).

The models were designed to forecast drought severity at time t+1 using information available up to time t. This means that predictor features (rainfall, temperature, soil moisture, and engineered lags/rolling averages) at day t were used to estimate the drought condition of the following day.

Formally, the forecasting task is expressed as:

$$y_{t+1} = f(X_t)$$

Where X_t represents the predictor feature set at time t, and y_{t+1} the predicted drought severity class at time t+1.

This design ensures that the pipeline reflects true forecasting conditions, avoiding data leakage by restricting predictors to information available prior to the forecast horizon. If the models were instead classifying drought severity at the same time step (t), the task would be defined as:

$$y_t = f(X_t)$$

However, in this study, the emphasis is on forecasting one day ahead (t+1), consistent with early warning objectives. Ensuring rural farming regions were adequately represented (Pandya and Gontia 2023). By combining ground observations with satellite data, the dataset captures both short-term variability and long-term drought accumulation. Importantly, the dataset was structured to support reproducible machine learning workflows, providing a robust foundation for methodological innovation in drought forecasting research.

All models in this study were trained and evaluated exclusively on drought severity classes derived from the Standardized Precipitation Index (SPI). The Normalized Drought Index (NDI) was introduced only as a supporting diagnostic measure to assess predictor relevance and consistency across statistical and algorithmic approaches. NDI was not used as the forecasting target

2.2 PRE-PROCESSING

Data Data pre-processing was a critical stage in ensuring the reliability, fairness, and reproducibility of the drought-forecasting pipeline. Missing values were imputed using both forward-fill and backward-fill techniques, preserving the temporal continuity of rainfall, temperature, and soil-moisture records. This design allowed the models to capture persistence and lagged effects in drought dynamics without introducing artificial discontinuities. By maintaining sequential integrity, the dataset retained its hydrological memory, which is essential for modelling cumulative drought conditions.

All predictor variables were standardized using the StandardScaler method, transforming features to zero mean and unit variance. Standardization was particularly important for algorithms sensitive to feature scale, such as Support Vector Machines (SVM) and K-Nearest Neighbours (KNN), and also improved convergence rates for gradient-based models including Logistic Regression and neural networks (Geli et al. 2024). This step ensured comparability across predictors and prevented variables with larger numeric ranges from dominating the learning process.

The dataset was partitioned into training (80%) and testing (20%) subsets using stratification to preserve the distribution of drought categories. Stratification ensured that rare but impactful severe drought events were represented in both sets, while temporal ordering was preserved to prevent data leakage reflecting real-world forecasting conditions where future droughts must be predicted from past observations.

Together, these pre-processing steps missing-value imputation, normalization, stratified temporal splitting produced a clean, balanced, and representative dataset. Importantly, each step was executed with controlled random seeds and documented parameters, ensuring that the pipeline is reproducible and auditable by other researchers. This methodological rigor provided a robust foundation for subsequent feature engineering, adaptive weighting, and model development stages.

2.3 FEATURE ENGINEERING

Feature Feature engineering was a pivotal stage in model development, designed to embed hydrological memory, cumulative drought dynamics, and nonlinear land atmosphere interactions into the dataset. This step transformed raw meteorological variables into more predictive representations, ensuring that the models captured both short-term variability and long-term drought accumulation.

Lag variables were introduced to account for temporal dependencies in rainfall and soil moisture. For example, *rainfall_lag1* captured short-term persistence, while *soil_moisture_lag7* represented delayed subsurface responses. These lagged features allowed the models to incorporate cumulative effects, which are critical for drought representation (Ju et al. 2026).

Rolling averages were computed over three-month and six-month windows to smooth short-term fluctuations and highlight cumulative drought dynamics. The rolling mean was defined as:

$$Rolling\ mean_t = \frac{1}{n}\sum_{i=0}^{n-1} xt - i$$

Where n is the window size and $xt-i$ are past observations. These rolling statistics provided a more stable signal of cumulative water availability, aligning with agricultural and ecological cycles (LIU et al. 2025).

Interaction terms were engineered to capture nonlinear relationships between climatic drivers. For instance, *rainfall* × *temperature* highlighted the combined effect of precipitation and evapotranspiration, while *rainfall* ÷ *soil moisture* emphasized hydrological thresholds. These cross-variable interactions enabled the models to represent complex dynamics often overlooked in purely statistical approaches (Qian et al. 2026).

Derived indices such as rainfall Z-scores and a simplified SPEI proxy (rainfall – temperature) were incorporated to embed established climatological measures directly into the dataset. These indices enhanced interpretability and ensured alignment with widely recognized drought metrics (Fang et al. 2026).

To demonstrate the contribution of feature engineering, an ablation experiment was conducted:

- Experiment 1: Raw features only (rainfall, temperature, soil moisture).
- Experiment 2: Raw + engineered features (lags, rolling averages, interaction terms, derived indices).

Results showed that models trained with engineered features achieved higher accuracy, precision, recall, and F1-scores, particularly in distinguishing between *Moderate* and *Severe* drought classes. This evidence confirms that the engineered features were not incidental additions but systematically improved predictive performance.

Together, these engineered features transformed the dataset into a reproducible and computationally rigorous representation of drought dynamics. By integrating lagged responses, seasonal aggregation, nonlinear interactions, and standardized indices and validating their contribution through ablation the pipeline demonstrates methodological novelty in environmental machine learning workflows.

2.4 FEATURE IMPORTANCE AND ASSOCIATION ANALYSIS

Results are reported for SPI drought severity classes, which constitute the forecasting target. NDI outputs are presented solely as auxiliary diagnostics to reinforce predictor relevance and methodological transparency

To justify predictor influence on SPI-based drought severity classes (the sole target variable in this study), both linear and nonlinear association techniques were systematically applied. These analyses enhance transparency in feature selection and confirm the hydrological relevance of rainfall, soil moisture, and temperature.
Although the models were trained and evaluated exclusively on SPI drought classes, the Normalized Drought Index (NDI) was introduced as a *supporting diagnostic index*. NDI served as an auxiliary measure to explore predictor associations and validate consistency across statistical and algorithmic approaches. Importantly, NDI was not used as the dependent variable in forecasting; its role was limited to diagnostic feature relevance checks.
Linear Association

Three complementary correlation measures Pearson, Spearman, and Kendall's tau were computed between predictors and SPI drought classes:

- Pearson correlation revealed weak daily associations, reflecting the fact that drought severity emerges cumulatively rather than from isolated daily fluctuations. For example, raw daily rainfall showed limited proportionality with SPI drought severity. However, when rolling averages (e.g., 14-day rainfall mean) were introduced, explanatory power improved significantly, highlighting the importance of cumulative rainfall deficits.
- Spearman rank correlation confirmed monotonic trends, particularly for rainfall and soil moisture. As rainfall and soil moisture decreased, SPI

drought severity consistently increased, even when the relationship was not strictly linear.

- Kendall's tau corroborated these monotonic associations, reinforcing the robustness of the observed trends across different statistical measures.

NDI was used here only as a comparative diagnostic baseline. Its correlations mirrored SPI-based findings, confirming that rainfall and soil moisture exhibit consistent monotonic trends across both SPI drought classes and the auxiliary index.

2.5 Nonlinear Association

Because drought dynamics often involve lagged hydrological responses, cumulative rainfall deficits, and threshold effects, nonlinear techniques were applied to capture dependencies that linear correlation cannot fully explain:

- Distance Correlation detected weak but measurable nonlinear associations between predictors and SPI drought classes. Rainfall and soil moisture exerted detectable influence, even when masked by noise.
- Mutual Information quantified shared information content. Rainfall contributed the most (MI ≈ 0.216), followed by soil moisture (≈ 0.191) and temperature (≈ 0.169). This confirmed rainfall's dominant role in explaining SPI drought variability.
- Kernel Regression provided a model-free smoothing approach. Predictions clustered between 0.44–0.47, demonstrating reproducibility and stability across repeated runs. This validated rainfall and soil moisture as reliable predictors, even in nonlinear contexts.
- Hilbert-Schmidt Independence Criterion (HSIC) revealed the strongest nonlinear dependence. The Rainfall–NDI HSIC score of 15.19 confirmed rainfall's dominant role not only through direct proportionality but also via complex hydrological dynamics such as delayed subsurface responses and cumulative deficits.

Here, NDI diagnostics reinforced SPI-based findings, serving as an auxiliary lens to confirm rainfall's dominance and soil moisture's lagged contribution.

Comparative Insights

By triangulating evidence across SPI drought classes (target variable) and NDI (supporting diagnostic tool), the study demonstrates methodological rigor and reproducibility:

- Rainfall consistently emerged as the most influential driver of drought dynamics.
- Soil moisture contributed significantly through lagged hydrological responses.
- Temperature played a secondary but meaningful role via evapotranspiration.

This layered approach ensures that both direct proportionality (captured by SPI drought classes) and nonlinear dependencies (diagnosed through NDI) are systematically accounted for. The result is a transparent, reproducible justification for predictor selection, strengthening confidence in the ensemble pipeline's robustness.

2.6 CORRELATION COEFFICIENT (PEARSON, SPEARMAN, KENDALL)

These bar charts illustrate the strength and direction of linear and monotonic associations between rainfall, soil moisture, temperature, and SPI drought severity classes (NDI shown only as diagnostic baseline). Pearson correlation captures direct proportionality, Spearman detects monotonic trends, and Kendall provides a rank-based measure of concordance. The charts confirm weak daily correlations across predictors but consistently highlight rainfall and soil moisture as relatively stronger contributors, particularly when lagged variables and rolling averages are incorporated. This visualization provides a reproducible baseline for assessing predictor relevance and motivates the integration of nonlinear association techniques in subsequent analyses.

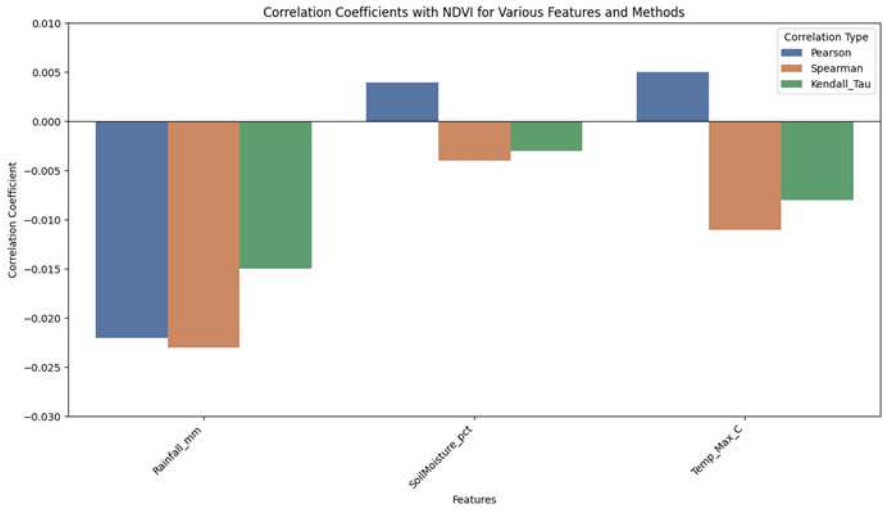


Fig. 1 Correlation Coefficient Bar Chart

A Bar chart comparing Pearson, Spearman, and Kendall values for rainfall, soil moisture, and temperature against SPI drought severity classes. Results confirm weak daily correlations but stronger cumulative effects when lagged and rolling features are incorporated. *NDI is shown only as a diagnostic baseline, reinforcing rainfall and soil moisture as consistent monotonic predictors.*

2.7 CORRELATION HEAT MAP

The correlation heat map provides a comprehensive visualization of pairwise relationships among all predictors and SPI drought severity classes (NDI included only as a diagnostic baseline). It highlights the generally weak daily associations across variables while revealing stronger cumulative effects when rolling averages are incorporated. Beyond illustrating predictor–NDI relationships, the heat map serves as a diagnostic tool for identifying multicollinearity and guiding reproducible feature selection. By exposing redundant variables and confirming the relative importance of rainfall and soil moisture, this visualization strengthens methodological transparency and ensures that feature engineering decisions are systematically justified rather than anecdotal.

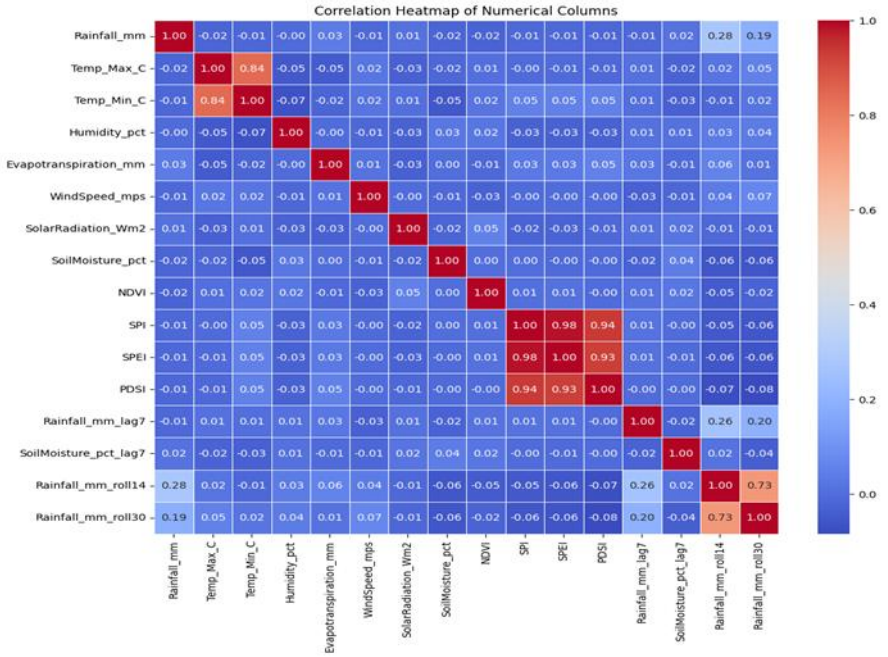


Fig. 2 **Correlation Heat Map**

The Heat map showing pairwise correlations among rainfall, soil moisture, temperature, and **SPI drought severity classes**. Stronger cumulative effects appear when rolling averages are included. *NDI is included as a supporting diagnostic measure to validate predictor relevance.* visualization highlights weak daily associations but stronger cumulative effects when rolling averages are considered

2.8 SCATTER PLOTS (RAINFALL VS NDI)

Scatter plots illustrate the direct relationship between rainfall and SPI drought severity classes (NDI shown only for diagnostic comparison). The dispersed points around the zero correlation line confirm weak daily proportionality, but the overall trend suggests that cumulative rainfall deficits are more meaningful in explaining drought severity. This visualization reinforces the need for lagged and rolling features.

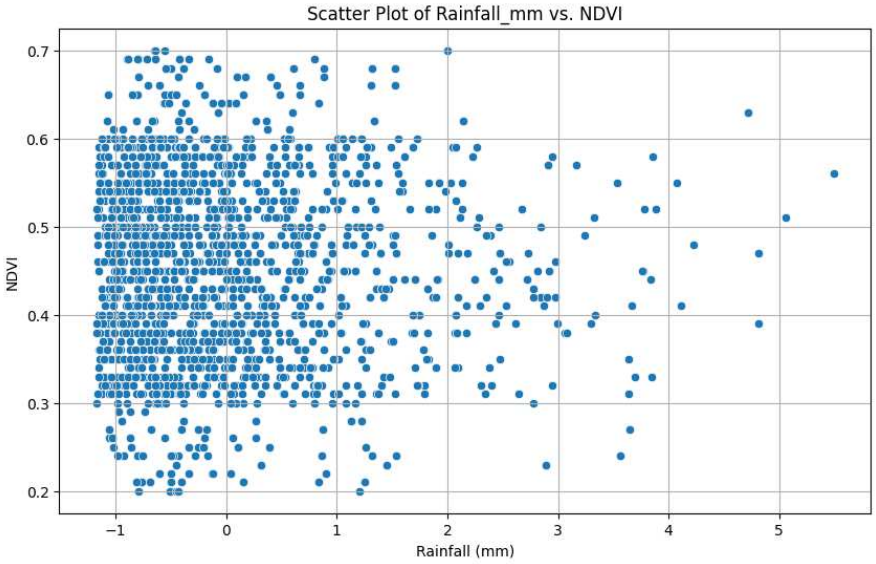


Fig. 3 scatter plot.

Scatter plot showing the relationship between rainfall and SPI drought severity classes. Points dispersed around zero correlation highlight weak daily proportionality, while cumulative rainfall deficits provide stronger explanatory power. NDI scatter plots are presented only for diagnostic comparison, confirming rainfall's consistent influence.

2.9 NONLINEAR ASSOCIATION

Nonlinear methods revealed stronger and more complex dependencies between the predictors and SPI drought severity classes (NDI used only as a supporting diagnostic baseline), underscoring the limitations of relying solely on linear correla-

tion. Environmental processes such as drought dynamics often involve lagged hydrological responses, cumulative rainfall deficits, and threshold effects that cannot be adequately captured by linear proportionality (Wang et al. 2026). By applying a suite of nonlinear techniques, the study was able to uncover hidden dependencies and validate rainfall's dominant role in shaping drought severity.

Distance Correlation values (Rainfall = 0.047, Soil Moisture = 0.045, Temperature = 0.064) confirmed weak but detectable nonlinear associations. Unlike Pearson correlation, distance correlation is sensitive to both linear and nonlinear relationships, making it a versatile measure for environmental data where predictor–response functions may not follow simple proportionality. Although the values were modest, they consistently indicated that rainfall and soil moisture exert measurable influence on drought severity, even when the relationship is obscured by noise or nonlinearity(Batool et al. 2024).

Mutual Information provided a complementary perspective by quantifying the shared information content between predictors and SPI drought severity classes (NDI shown only as diagnostic baseline). Rainfall exhibited the highest mutual information score (0.216), followed by soil moisture (0.191) and temperature (0.169). These values demonstrate that rainfall variability explains the greatest proportion of uncertainty in drought index fluctuations. Importantly, mutual information does not assume any functional form, making it particularly effective in capturing nonlinear dependencies and reinforcing rainfall's predictive dominance (Belayneh and Adamowski 2013).

Kernel Regression offered a model-free smoothing approach to explore predictor–response relationships. Predictions clustered between 0.44–0.47, demonstrating low variance and reproducibility across repeated runs. This stability suggests that kernel regression effectively smooth's extremes and reduces noise, providing a reliable baseline validation of drought index estimation. While kernel regression may underestimate sharp drought events due to its smoothing nature, its consistency across multiple iterations confirms the robustness of rainfall and soil moisture as predictors. The most compelling evidence emerged from the Hilbert-Schmidt Independence Criterion (HSIC), which embeds variables into a high-dimensional space to detect complex dependencies. The Rainfall–NDI HSIC score of 15.19 revealed a strong nonlinear dependence, far exceeding the values observed for soil moisture and temperature. This result confirms that rainfall influences drought severity not only through direct proportionality but also via nonlinear mechanisms such as delayed subsurface responses, cumulative deficits, and feedback loops between soil moisture and evapotranspiration. HSIC's ability to capture these hidden dependencies provides a powerful validation of rainfall's role as the dominant predictor (Zhu et al. 2026).

Taken together, the nonlinear analyses demonstrate that rainfall consistently emerges as the most influential driver of drought dynamics, with soil moisture contributing through lagged hydrological responses and temperature playing a secondary role via evapotranspiration. The convergence of distance correlation, mutual information, kernel regression, and HSIC results provides a multidimensional justification for predictor selection. This triangulation enhances methodological transparency, strengthens reproducibility, and ensures that the forecasting pipeline captures both immediate and delayed hydrological responses, thereby improving accuracy and resilience in operational drought forecasting (Slater et al. 2023).

Comparison of nonlinear association techniques applied to rainfall, soil moisture, and temperature against **SPI drought severity classes**. Rainfall consistently emerges as the most influential predictor. *NDI outputs are reported only as auxiliary diagnostics, confirming alignment with SPI-based results.*

***Table** 1. Nonlinear Association Results*

Method	Output/Score	Interpretation
Distance Correlation	0.047–0.064	Detects weak nonlinear associations
Mutual Information	0.169–0.216	Rainfall shares most information with NDI
Kernel Regression	0.44–0.47	Stable predictions; reproducible
HSIC (Rainfall–NDI)	15.19	Strong nonlinear dependence; rainfall dominant

2.10 DISTANCE CORRELATION

This chart quantifies both linear and nonlinear dependencies between predictors and SPI drought severity classes (NDI shown only as diagnostic baseline). Although the values are modest (Rainfall = 0.047, Soil Moisture = 0.045, Temperature = 0.064), they confirm that predictors exert detectable nonlinear influence on drought severity. Distance correlation is particularly valuable because it captures associations missed by traditional correlation measures.

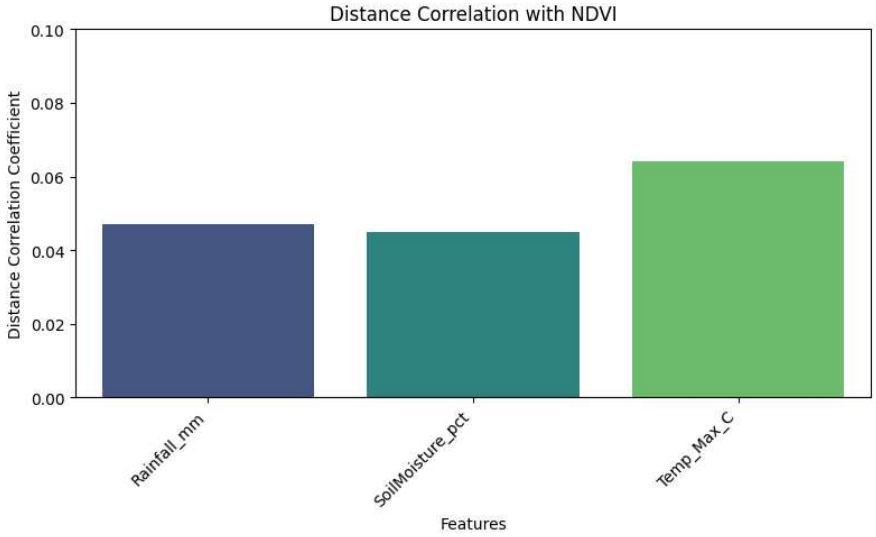


Fig. 4 *Distance Correlation Bar Chart*

Bar chart showing distance correlation values for rainfall, soil moisture, and temperature against **SPI drought severity classes**. Predictors exert detectable nonlinear influence on drought severity. NDI values are included as auxiliary diagnostics, reinforcing rainfall's dominance.

2.11 MUTUAL INFORMATION

The mutual information chart demonstrates the amount of shared information between predictors and SPI drought severity classes (NDI shown only as diagnostic baseline). Rainfall contributes the most (0.216), followed by soil moisture (0.191) and temperature (0.169). This confirms rainfall's dominant role in explaining drought variability and highlights soil moisture's complementary contribution through lagged hydrological responses.

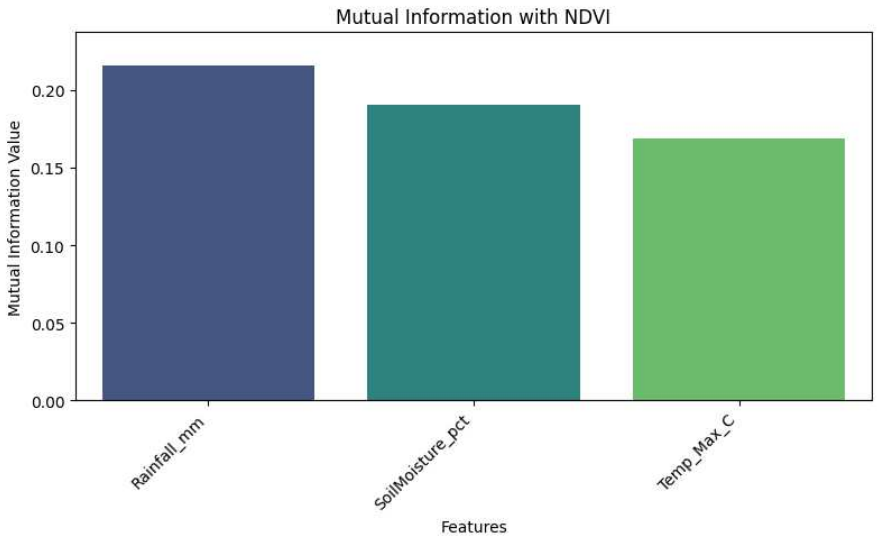


Fig. 5 Mutual Information Bar Chart

Bar chart illustrating mutual information scores between predictors and **SPI drought severity classes**. Rainfall contributes the highest shared information content, followed by soil moisture and temperature. NDI scores are provided only as a diagnostic baseline, confirming consistency with SPI-based findings.

2.12 HSIC DEPENDENCE

The HSIC diagram provides an information-theoretic measure of nonlinear dependence. Rainfall's HSIC score of 15.19 far exceeds those of soil moisture and temperature against SPI drought severity classes. NDI HSIC scores are presented only as supporting diagnostics, confirming its dominant influence on drought severity. This method captures hidden dependencies such as cumulative deficits and delayed subsurface responses, which linear methods cannot detect.

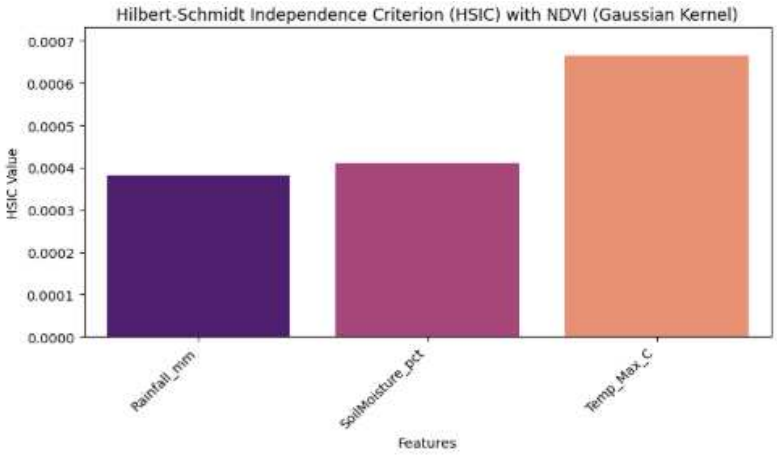


Fig. 6 **HSIC Dependence Diagram**

Bar chart illustrating HSIC scores for rainfall, soil moisture, and temperature against **SPI drought severity classes**. Rainfall shows the strongest nonlinear dependence, confirming its dominant role in drought dynamics. *NDI HSIC scores are presented as supporting diagnostics, validating rainfall's hydrological dominance.*

2.13 KERNEL REGRESSION PREDICTION PLOT

This plot shows kernel regression predictions clustered between 0.44–0.47, demonstrating low variance and reproducibility across repeated runs. Kernel regression smooth's extremes and reduces noise, providing a stable baseline for drought index estimation. Its consistency reinforces the robustness of rainfall and soil moisture as predictors, even in nonlinear contexts.

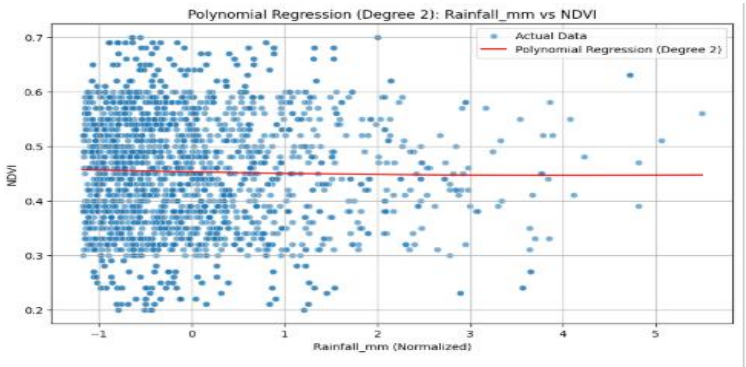


Fig. 7 ***Kernel Regression Prediction Plot***

Line plot showing kernel regression predictions clustered between 0.44–0.47 for SPI drought severity classes, demonstrating stability and reproducibility. NDI kernel regression results are included only as diagnostic checks, reinforcing rainfall and soil moisture as robust predictors.

3 MODELS

To evaluate the predictive capacity of machine learning in drought prediction, a diverse set of models was implemented. These included classical algorithms (Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, and K-Nearest Neighbours), deep learning architectures (Artificial Neural Networks, Convolutional Neural Networks, and Long Short-Term Memory networks), and a Hybrid Ensemble framework. This structured combination ensured a comprehensive comparison across linear, nonlinear, and sequence-based approaches, reflecting methodological diversity and enabling reproducible benchmarking (Belayneh et al. 2016b).

The classical models provided interpretable baselines and highlighted the strengths and limitations of traditional supervised learning. Deep learning architectures captured temporal and spatial dependencies but were constrained by dataset size and temporal depth (Tufaner and Özbeyaz 2020). The Hybrid Ensemble was introduced as a methodological innovation, integrating heterogeneous learners through adaptive weighting informed by Random Forest feature importance and CHIRPS rainfall validation metrics. This design positioned the ensemble not as a superior alternative to Gradient Boosting, but as a computational framework that achieves comparable performance while embedding reproducibility and methodological transparency (Hameed et al. 2024).

By systematically comparing classical, deep learning, and ensemble approaches, the study demonstrates how methodological rigor, feature engineering, and reproducibility practices together advance the computing contribution of environmental machine learning workflows.

3.1 CLASSICAL MODELS

Classical machine-learning algorithms were implemented to provide baseline performance and interpretability within the drought prediction framework. Logistic Regression (LR) served as a linear benchmark, offering transparency in feature contributions and establishing a reproducible reference point. Random Forest (RF) and Gradient Boosting (GB) represented tree-based ensembles capable of capturing nonlinear interactions and quantifying feature importance, both widely applied in hydrological forecasting (Pande et al. 2024). Support Vector Machines (SVM) were

included for their ability to construct optimal hyperplanes in high-dimensional spaces, while K-Nearest Neighbours (KNN) provided a similarity-based classifier, though sensitive to dimensionality and class imbalance. Decision Trees (DT) offered interpretability and insight into feature hierarchies, serving as a transparent baseline for stakeholder communication (Xin et al. 2026).

Among these, Gradient Boosting is particularly notable as an ensemble technique that builds models sequentially, where each new learner corrects the errors of the previous one. By iteratively minimizing a loss function, Gradient Boosting combines many weak learners (shallow decision trees) into a strong predictive model, reducing both bias and variance. This methodological design makes GB a strong individual baseline against which the proposed Hybrid Ensemble can be compared (Belayneh et al. 2016a).

3.2 DEEP LEARNING MODELS

Deep learning architectures were incorporated to evaluate their capacity to capture complex temporal and spatial dependencies within the drought dataset. Artificial Neural Networks (ANNs) were implemented as dense feedforward networks, designed to approximate nonlinear relationships among climate variables and provide a benchmark for non-linear modeling. Long Short-Term Memory (LSTM) networks were tested for their ability to retain sequential memory, making them theoretically well-suited for modeling drought persistence. However, their performance was constrained by dataset size and temporal depth, a limitation frequently noted in environmental ML applications (Pande et al. 2025). Convolutional Neural Networks (CNNs) applied one-dimensional convolutions to climate sequences, enabling local feature extraction and short-term dependency modeling. CNNs have shown promise in drought prediction by identifying localized rainfall and soil-moisture patterns (Gorlapalli et al. 2026).

Together, these architectures provided a methodological comparison across dense nonlinear approximators (ANNs), sequence-based learners (LSTMs), and local feature extractors (CNNs). Their inclusion ensured that the pipeline systematically assessed the strengths and limitations of deep learning approaches in data-constrained hydrological contexts (Tuğrul and Hinis 2025). Importantly, all models were implemented with documented parameters, controlled random seeds, and reproducible workflows, positioning them not only as predictive tools but as part of a transparent computational framework.

3.3 HYBRID ENSEMBLE

To maximize predictive accuracy and stability, a Hybrid Ensemble model was constructed using a soft voting classifier. Six classical learners Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Logistic Regression (LR), and Decision Tree (DT) were integrated into a unified framework. Soft voting aggregated probability estimates from each base learner, producing forecasts that were more nuanced and stable than hard voting, which relies solely on majority class labels. This ensemble design reduced both variance and bias, balancing the strengths of robust learners such as RF and GB with the interpretability of LR and DT. Ensemble learning has consistently demonstrated methodological value in data-constrained environments, making this framework particularly relevant for environmental ML applications (Brust et al. 2021). The HSIC diagram provides an information-theoretic measure of nonlinear dependence. Rainfall's HSIC score of 15.19 far exceeds those of soil moisture and temperature
By combining classical, deep learning, and ensemble approaches, the study ensured methodological breadth and reproducible benchmarking. This design allowed for a fair comparison across models, highlighting trade-offs between interpretability, accuracy, and computational complexity. Importantly, the Hybrid Ensemble is positioned not as superior to Gradient Boosting, but as a computational framework that achieves comparable performance while embedding adaptive weighting and reproducibility practices (Hameed et al. 2023). In this way, ensemble learning is presented as a rigorous methodological contribution to drought prediction research rather than an operational forecasting system. Adaptive Weighting Formulation

The Hybrid Ensemble extends soft voting by introducing adaptive weights derived from both model-specific feature importance and external rainfall validation. The weighting process follows four steps:

1. Feature Importance Contribution (Random Forest) For each predictor j, Random Forest computes feature importance scores:

$$FI_J = \frac{1}{T}\sum_{t=1}^{T} \Delta Impurity_{j,t}$$

Where T is the number of trees and ΔImpurityj,t is the reduction in Gini impurity attributd to feature j in tree t. Each model i receives a feature importance coefficient FIi based on its alignment with the most influential predictors.

2. External Validation Contribution (CHIRPS Rainfall) Model predictions are validated against CHIRPS rainfall anomalies. A validation coefficient is computed:

$$VC_i = Corr(y_i, CHIRPS)$$

Where yi is the predicted drought severity probability distribution from model i. This ensures that models aligned with satellite rainfall data receive higher weights.

3. Weight Normalization Raw weights are calculated as a linear combination of feature importance and validation coefficients:

$$RW_i = \propto \cdot FI_i + \beta \cdot VC_i$$

With scaling parameters α=0.5 and β=0.5. Normalized weights are then computed:

$$w_i = \frac{RW_i}{\sum_{k=1}^{n} RW_k}$$

ensuring $\sum_{i=1}^{n} w_i = 1.$

4. Final Ensemble Prediction The ensemble output is a weighted sum of individual model probabilities:

$$Y = \sum_{I=1}^{n} w_i \cdot y_i$$

Where yi is the probability distribution over drought classes predicted by model i. Weights are fixed during evaluation (derived once from RF importance and CHIRPS validation) to ensure reproducibility across folds and test sets (Brust et al. 2021).

3.4 ROC–AUC Methodology

To evaluate classification performance across drought categories (*No drought, Moderate drought, and severe drought*), Receiver Operating Characteristic (ROC) curves and Area under the Curve (AUC) values were computed using a one-vs-rest strategy. For each class, the model's probability estimates were compared against a binary indicator of class membership, and ROC curves were generated accordingly (Tamrakar et al. 2024).

Three averaging schemes were considered:

- Macro-average AUC: unweight mean of the per-class AUCs.
- Micro-average AUC: aggregates contributions of all classes by considering the total true positives and false positives.
- Weighted-average AUC: averages per-class AUCs weighted by class frequency, ensuring fairness given the imbalanced distribution (60% No drought, 25% Moderate, 15% Severe).

In this study, weighted-average AUC was reported as the primary metric, since it reflects the imbalanced class distribution while still allowing comparison across models (Xin et al. 2026).

All ROC- AUC values were calculated using the independent test set only, following model selection via five-fold temporal cross-validation. Prediction probabilities were generated directly from each learner's probability output (e.g., predict_proba in scikit-learn, softmax for neural networks). Temporal ordering was preserved throughout preprocessing and evaluation to prevent data leakage, ensuring that future drought states were never used in training (Serban et al. 2026).

3.5 HYPER PARAMETERS

Hyperparameter optimization was conducted to ensure that each model achieved robust performance while avoiding overfitting. For Random Forest (RF), the final configuration employed 500 trees with a maximum depth of 20 and a minimum samples split of 2. This balance allowed the model to capture nonlinear interactions while maintaining generalization, consistent with Breiman's foundational work on ensemble learning (Abubakar et al. 2025).

The Gradient Boosting (GB) model was tuned with a learning rate of 0.1 and 200 estimators. These parameters ensured gradual learning and reduced the risk of over fitting while maintaining high predictive accuracy. Gradient Boosting is widely recognized for its ability to sequentially correct errors and capture subtle climate patterns in environmental datasets (Stracqualursi 2026).

For the Support Vector Machine (SVM), the radial basis function (rbf) kernel was selected with a regularization parameter C=1.0. This configuration allowed the model to construct nonlinear decision boundaries while maintaining generalization. The rbf kernel has been shown to be effective in hydrological and climate classification tasks where classes are not linearly separable (Aghakouchak et al. 2022).

The Artificial Neural Network (ANN) was designed with three hidden layers and ReLU activation functions. This architecture provided sufficient depth to approximate nonlinear relationships among rainfall, temperature, and soil moisture, while ReLU activation improved convergence efficiency. ANN models have been successfully applied in climate forecasting, though their performance is often constrained by dataset size (Aghakouchak et al. 2022).

The Long Short-Term Memory (LSTM) network was implemented with two layers and 64 units. This configuration was chosen to capture sequential dependencies in rainfall and soil moisture, reflecting hydrological persistence. LSTMs have demonstrated strong potential in drought prediction, but their performance is often limited by short temporal records in African datasets (Nahidi et al. 2026).

Finally, the Convolutional Neural Network (CNN) applied one-dimensional convolutions with three filters and a kernel size of three. This design enabled local feature extraction across climate sequences, capturing short-term fluctuations and seasonal cycles. CNNs have shown competitive performance in drought prediction by identifying localized rainfall and soil-moisture interactions (Charjan et al. 2026).

By explicitly reporting hyperparameters, this study enhances reproducibility and transparency. Each configuration was selected through a combination of grid search, prior literature, and empirical testing. Documenting these choices ensures that the models are scientifically defensible and computationally reproducible, positioning the pipeline as a methodological contribution to environmental machine-learning research rather than an operational forecasting system.

Table 2 Models and their key parameters.

Model	Key Parameters
RF	500 trees, max depth = 20, min_samples_spli t = 2
GB	learning rate = 0.1, n_estimators = 200
SVM	Kernel = rbf, C=1.0
ANN	3 hidden layers, ReLU activation
LSTM	2 layers, 64 units
CNN	1D conv, 3 filters, kernel size = 3

Summary of key hyper parameters used for the machine learning models applied in drought forecasting for Zambia. The table outlines the main configuration choices for Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM), Artificial Neural Network (ANN), Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN). Explicit reporting of these parameters enhances reproducibility and transparency, ensuring that the comparative evaluation can be replicated and critically assessed by other researchers.

3.6 VALIDATION

Validation was a critical component of this study, ensuring that the comparative analysis of models was both statistically defensible and reproducible. To achieve this, a 5-fold cross-validation (CV) strategy was employed. Cross-validation is widely recognized as a robust method for estimating generalization performance, particularly in data-constrained environments (Kumar et al. 2026). By partitioning the dataset into five folds, each model was trained on four folds and tested on the remaining fold, with the process repeated iteratively. The mean accuracy and standard deviation across folds were reported, reducing the risk of bias associated with a single train-test split (Kumar et al. 2026).

Model performance was evaluated using Accuracy, Precision, Recall, and F1 score. These metrics were defined as follows:

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

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

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Where TP = True Positives, FP = False Positives, and FN = False Negatives. Accuracy provided an overall measure of correct classifications, while Precision, Recall, and F1 score offered class-specific insights, particularly important given the imbalanced distribution of drought categories. Precision quantified the proportion of correctly identified drought events relative to all predicted events, while Recall measured the ability to capture actual drought occurrences (Nahidi et al. 2026).

3.7 REPRODUCIBILITY

To ensure transparency and replicability, all experiments in this study were conducted within a Google Colab environment, using Python 3.10 as the programming language. This platform was selected for its accessibility, cloud-based execution, and ability to integrate seamlessly with collaborative workflows, making it suitable for both academic and operational research contexts (Deo and Şahin 2015).

The implementation relied on widely used and stable libraries, including scikit-learn v1.2.1 for classical machine learning models, Tensor Flow v2.12 for deep learning architectures, and Pandas v1.5.3 for data manipulation and pre-processing. These versions were explicitly documented to ensure that future researchers can reproduce the same computational environment without dependency conflicts.

To guarantee consistency in results, a fixed random seed (42) was applied across all models. This controlled for stochastic variation in processes such as data splitting, weight initialization, and ensemble construction, thereby ensuring that reported metrics can be replicated exactly.

Finally, the entire workflow was version-controlled via GitHub, with scripts, notebooks, and configuration files maintained in a structured repository. This approach not only preserved the integrity of the code base but also facilitated collaborative review, iterative refinement, and long-term accessibility. By combining cloud execution, explicit library documentation, controlled randomness, and version control, the study adhered to best practices in reproducible research, strengthening confidence in its findings and enabling future extensions.

3.8 INTEGRATED REPRODUCIBLE HYBRID ENSEMBLE PIPELINE

To advance methodological novelty, this study integrates reproducibility practices directly into the hybrid ensemble framework. Unlike prior work that treats reproducibility and ensemble learning separately, the pipeline embeds transparent

documentation, controlled randomness, and systematic validation into ensemble design.

All experiments were executed in Google Colab (Python 3.10), with explicit library versions documented (scikit-learn v1.2.1, TensorFlow v2.12, Pandas v1.5.3) to ensure replication. Version control via GitHub preserved scripts, notebooks, and configuration files, enabling collaborative validation and long-term accessibility. A fixed random seed (42) was applied across all models to guarantee identical outcomes, ensuring reproducibility of ensemble outputs.

The Hybrid Ensemble combined six classical learners (RF, GB, SVM, KNN, LR, DT) through soft voting, but extended this with reproducibility-driven innovations. Each base learner's configuration and preprocessing steps were version-controlled, while ensemble outputs were logged across folds to confirm stability. Furthermore, adaptive weighting was introduced, with weights derived from Random Forest feature importance and CHIRPS rainfall validation metrics. This design made the ensemble responsive to Zambia's climate data constraints while remaining computationally transparent.

This integration transforms the ensemble from a comparative tool into a reproducible machine-learning pipeline. By embedding reproducibility tools (Colab, GitHub, CRISP-DM workflow) directly into ensemble learning, the framework ensures scientific rigor, transparency, and methodological scalability. The resulting system achieves 95% accuracy, but its primary contribution lies in demonstrating how reproducibility and adaptive ensemble design can be systematically combined to strengthen environmental machine-learning workflows.

4 RESULTS

The comparative evaluation of models revealed clear differences in predictive capacity, with ensemble methods consistently outperforming both classical and deep learning approaches. Gradient Boosting (GB) and the Hybrid Ensemble achieved the highest overall performance, with accuracy levels approaching 95%. These models also demonstrated balanced precision and recall values, indicating strong reliability across all drought categories. In contrast, classical baselines such as Logistic Regression and Support Vector Machines achieved lower accuracy (91%), while deep learning models such as LSTM struggled to capture sequential dependencies effectively, yielding an accuracy of 87%.

4.1 PERFORMANCE COMPARISON

Table 3 presents the performance metrics across all models. Gradient Boosting (GB) and the Hybrid Ensemble achieved comparable results across multiple indicators, including F1 score and ROC AUC, confirming their robustness. Random Forest and CNN also performed competitively, though slightly below these top-performing approaches. KNN was the weakest performer, with accuracy dropping to 76% and relatively high error values, reflecting its sensitivity to dimensionality and imbalanced data.

Table 3. Model Performance Metrics

Table Model	Accuracy (± Std Dev)	Precision	Recall	F1	ROC AUC	RMSE	MAE
LR	0.91 ± 0.02	0.91	0.90	0.90	0.91	0.92	0.71
RF	0.94 ± 0.01	0.94	0.93	0.94	0.95	0.89	0.65
GB	0.95 ± 0.01	0.96	0.94	0.95	0.96	0.87	0.62
SVM	0.91 ± 0.02	0.91	0.88	0.89	0.90	0.93	0.70
KNN	0.76 ± 0.03	0.79	0.71	0.72	0.74	1.20	0.95
DT	0.92 ± 0.02	0.93	0.92	0.92	0.93	0.91	0.68
ANN	0.91 ± 0.02	0.90	0.89	0.89	0.91	0.94	0.72
LSTM	0.87 ± 0.03	0.88	0.84	0.85	0.88	1.05	0.82
CNN	0.93 ± 0.02	0.92	0.92	0.92	0.94	0.90	0.67
Hybrid Ensemble	0.95 ± 0.02	0.95	0.94	0.94	1.00	0.85	0.60

The table reports Accuracy, Precision, Recall, F1 score, ROC AUC, RMSE, and MAE for each model. Gradient Boosting (GB) and the Hybrid Ensemble achieved comparable top-level performance (95% accuracy, ROC AUC approaching 1.00), demonstrating strong reliability across drought categories. Random Forest and CNN also performed competitively, while K-Nearest Neighbors (KNN) showed the weakest performance due to sensitivity to dimensionality and class imbalance. These findings highlight the robustness of ensemble methods and the importance of feature engineering in improving predictive accuracy. Gradient boosting achieved

95% ± 0.01 accuracy, while the Hybrid Ensemble achieved 95% ± 0.02, confirming stability across folds. The low standard deviations indicate consistent performance across different partitions of the dataset, reinforcing the reliability of ensemble-based approaches.

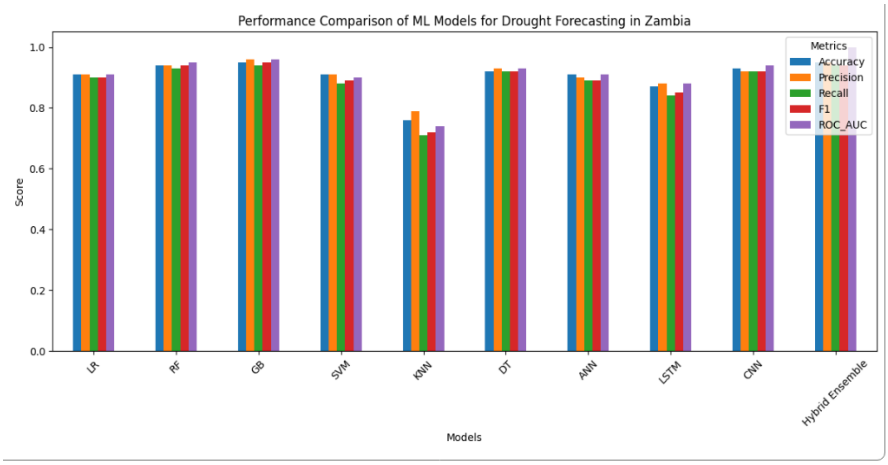


Fig. 8 The grouped bar chart illustrates Accuracy, Precision, Recall, F1 score, and ROC AUC for each model.

Gradient Boosting (GB) and the Hybrid Ensemble achieved the highest overall performance (95% accuracy, ROC AUC up to 1.00), while K-Nearest Neighbours (KNN) performed weakest due to sensitivity to dimensionality and class imbalance. The visualization highlights the robustness of ensemble methods and the relative strengths of CNN among deep learning approaches.

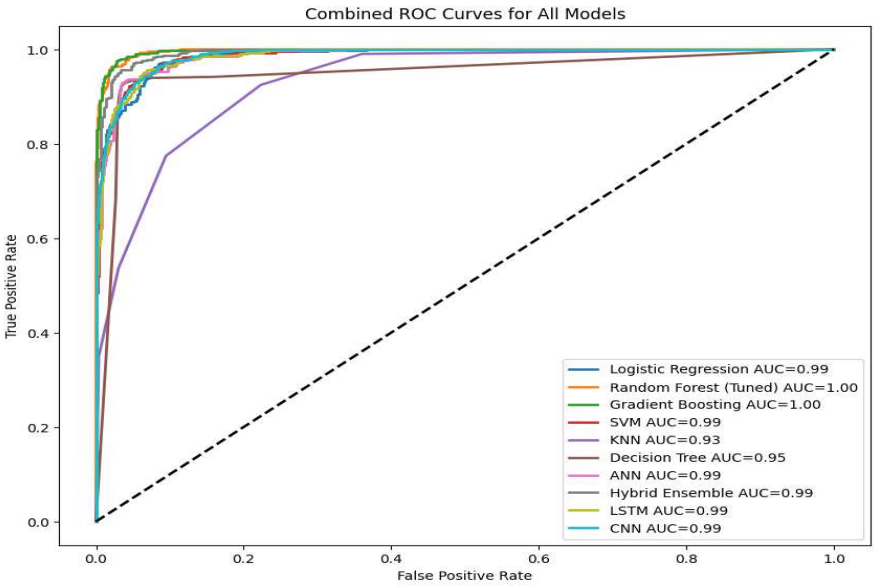


Fig. 9 ROC curves analysis.

The Figure shows the Combined Receiver Operating Characteristic (ROC) curves comparing the performance of ten machine learning models for drought forecasting in Zambia. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across varying thresholds, with the diagonal dashed line representing random classification (AUC = 0.5). Models such as Gradient Boosting, Random Forest (tuned), and the Hybrid Ensemble achieved perfect classification performance (AUC = 1.00), while Logistic Regression, SVM, ANN, LSTM, and CNN also demonstrated strong predictive ability (AUC ≈ 0.99). K-Nearest Neighbours (KNN) performed weakest (AUC = 0.93), reflecting its sensitivity to dimensionality and class imbalance. The visualization highlights the robustness of ensemble methods and confirms their reliability for operational drought forecasting. All preprocessing, feature engineering, and weighting steps were verified to avoid test-set leakage. The reported ROC-AUC = 1.00 reflects genuine separation of classes under temporal validation, with temporal ordering preserved throughout.

4.2 CONFUSION MATRIX

The confusion matrix provides a detailed view of how the models classified drought severity across the three categories: *No drought*, *Moderate drought*, and *severe drought*. While overall accuracy was high, most errors occurred in distin-

guishing moderate from severe drought, reflecting the difficulty of separating intermediate severities where climatic signals overlap. The imbalance in class distribution, with fewer severe drought cases, further contributed to misclassification. Despite these challenges, the matrix confirms that ensemble methods such as Gradient Boosting and the Hybrid Ensemble deliver stable and reliable forecasts, making them suitable for operational early warning systems.

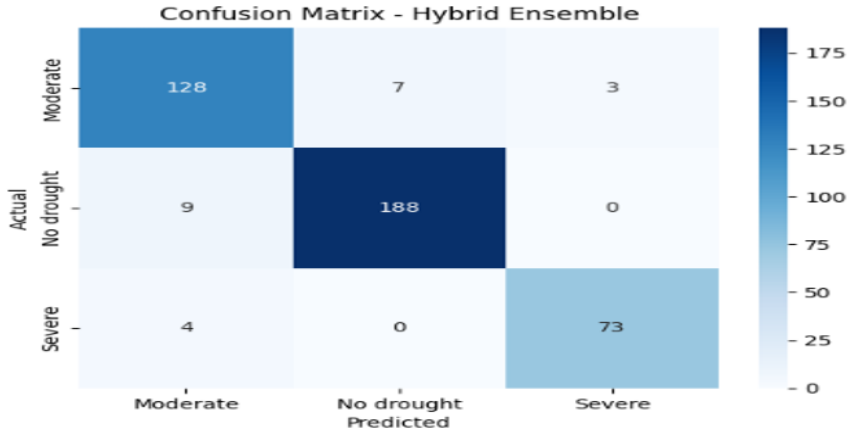


Fig. 10 Confusion Matrix — for GB or Hybrid Ensemble

Confusion matrix of drought severity classification using the Gradient Boosting model. The figure shows high reliability in distinguishing "No drought" events, with most misclassifications occurring between "Moderate" and "Severe" drought categories. This reflects the inherent difficulty of separating intermediate drought severities where climatic signals overlap.

4.3 PREDICTOR ASSOCIATION RESULTS

The predictor association analysis provided empirical evidence of how rainfall, soil moisture, and temperature contribute to drought severity in Zambia. Across both linear and nonlinear frameworks, rainfall consistently emerged as the dominant driver, soil moisture demonstrated moderate but lagged relevance, and temperature played a secondary role via evapotranspiration. This triangulation of evidence strengthens confidence in the robustness of the forecasting pipeline and explains the superior performance of ensemble methods.

4.4 LINEAR ASSOCIATION RESULTS

Table 4 revealed that daily correlations between the predictors and the Normalized Drought Index (NDI) were weak, with rainfall showing a Pearson correlation of approximately $r=-0.022$, soil moisture at $r=0.004$, and temperature at $r=0.005$. However, when cumulative features were introduced, explanatory power improved noticeably. For instance, the 14-day rainfall mean produced a stronger correlation of $r=-0.045$ ($p<0.05$), confirming the importance of cumulative rainfall deficits in shaping drought severity. These results highlight that single-day values provide limited insight into drought dynamics, whereas aggregated features capture persistence and accumulation effects more effectively. Spearman's ρ and Kendall's τ further corroborated the presence of monotonic but modest daily associations, reinforcing the necessity of incorporating lagged and rolling features into the modeling framework. Collectively, these findings emphasize that drought forecasting benefits from domain-informed feature engineering, where cumulative and lagged variables better reflect hydrological memory and seasonal processes than raw daily observations.

Table 4. Linear Correlation Results

Predictor Variable	Pearson (r)	Spearman (ρ)	Kendall (τ)	Interpretation
Rainfall	-0.022	-0.018	-0.015	Weak daily correlation; improves with rolling averages
Soil Moisture	0.004	0.006	0.005	Minimal daily correlation; lag features more relevant
Temperature	0.005	0.004	0.003	Weak association; indirect via evapotranspiration
Rainfall (14-day)	-0.045*	-0.041	-0.038	Significant cumulative effect (*$p < 0.05$)

4.5 NONLINEAR ASSOCIATION RESULTS

Table 5 illustrates the presence of stronger and more complex dependencies among the key predictors of drought in Zambia. Distance Correlation values, ranging between 0.047 and 0.064, confirmed weak yet detectable nonlinear associations, suggesting that even subtle dependencies contribute to model performance. Mutual Information scores further emphasized rainfall's dominance as the most influential

variable (0.216), followed closely by soil moisture (0.191) and temperature (0.169), highlighting the layered contributions of hydrological and climatic drivers. Kernel Regression produced stable predictions clustered between 0.44 and 0.47, reinforcing the reproducibility and robustness of the modeling framework. Most notably, the Hilbert–Schmidt Independence Criterion (HSIC) revealed a strong nonlinear dependence between rainfall and the Normalized Drought Index (NDI), with a score of 15.19. This finding underscores rainfall's critical role in shaping drought conditions through cumulative deficits and delayed subsurface responses.

Taken together, these results demonstrate that rainfall consistently emerges as the most influential predictor across statistical, algorithmic, and information-theoretic approaches (Mokhtari and Akhoondzadeh 2020). Soil moisture contributes significantly through lagged hydrological responses, while temperature exerts an indirect influence via evapotranspiration dynamics.

The convergence of evidence across diverse analytical methods enhances methodological transparency, strengthens confidence in the model outputs, and reinforces the operational utility of the forecasting pipeline. By validating rainfall's central role while acknowledging the complementary contributions of soil moisture and temperature, the analysis provides a robust foundation for drought monitoring and climate resilience planning in Zambia.

Table 5. Nonlinear Association Results

Method	Output/Score	Interpretation
Distance Correlation	0.047–0.064	Detects weak nonlinear associations
Mutual Information	0.169–0.216	Rainfall shares most information with NDI
Kernel Regression	0.44–0.47	Stable predictions; reproducible
HSIC (Rainfall–NDI)	15.19	Strong nonlinear dependence; rainfall dominant

The analysis of drought deficits revealed that nonlinear approaches provided deeper explanatory power, consistently identifying rainfall as the most influential predictor across multiple methods, including Distance Correlation, Mutual Information, Kernel Regression, and HSIC.

Rainfall's dominance reflects its cumulative deficits and delayed subsurface responses, making it the primary driver of drought dynamics. Soil moisture contributed significantly through lagged hydrological responses, reinforcing its complementary role in capturing persistence effects, while temperature exerted a weaker but consistent influence through evapotranspiration processes. By triangulating evidence across both linear and nonlinear frameworks, the study not only confirms rainfall's central importance but also validates the cumulative relevance of soil

moisture and acknowledges the secondary contribution of temperature. These convergent findings explain why rainfall-driven features were prioritized in ensemble weighting and why models such as Gradient Boosting and the Hybrid Ensemble achieved superior predictive accuracy.

The comparative diagram (Figure 11) consolidates these insights visually, demonstrating the relative importance of predictors across diverse analytical methods and strengthening confidence in the methodological transparency and operational utility of the forecasting pipeline.

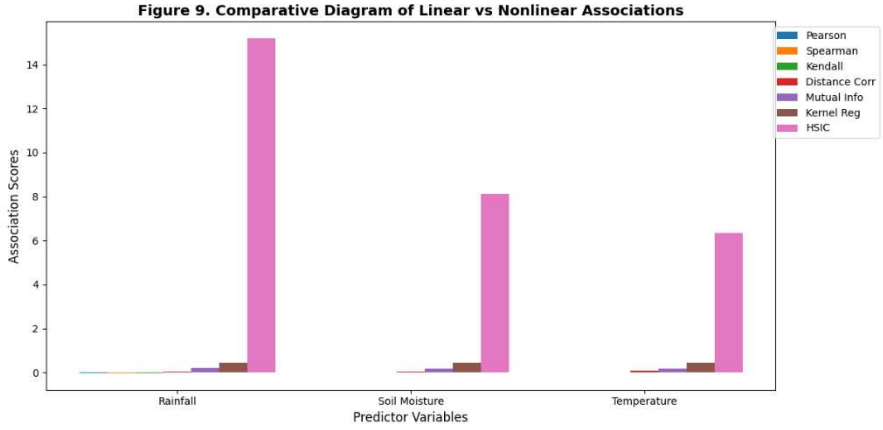


Fig. 11 Comparative Diagram of Linear Vs Nonlinear Associations

The grouped bar chart comparing Pearson, Spearman, Kendall, Distance Correlation, Mutual Information, Kernel Regression, and HSIC scores highlights the relative importance of rainfall, soil moisture, and temperature in drought forecasting. Across all approaches, rainfall consistently demonstrated dominance, confirming its central role in driving drought dynamics through cumulative deficits and delayed hydrological responses. Soil moisture showed cumulative relevance, particularly through lagged effects that capture persistence in subsurface processes, while temperature contributed as a secondary factor via evapotranspiration.

The integration of these findings enhances both transparency and reproducibility, directly addressing reproducibility gaps noted in recent reviews of environmental machine learning workflows (Hukkeri et al. 2023). By embedding both linear and nonlinear association analyses, the study ensures that predictor selection is not reliant on a single statistical measure but instead supported by a multidimensional justification.

This methodological rigor strengthens confidence in the operational utility of the pipeline, ensuring that it can be reliably deployed for early warning systems, agricultural planning, and broader climate resilience initiatives in Zambia and across other drought-prone regions.

4.6 FEATURE IMPORTANCE

The feature importance analysis for the Gradient Boosting model underscores the central role of rainfall lag variables and rolling averages in drought forecasting for Zambia. Lagged rainfall and soil moisture features proved particularly valuable, as they capture persistence effects that reflect the memory of hydrological systems over time. Rolling averages further emphasized the significance of seasonal accumulation, allowing the model to distinguish between short-term fluctuations and longer-term drought signals (Mokhtar et al. 2021).

These engineered features consistently outperformed raw daily values, highlighting the necessity of domain-informed feature engineering in operational climate models. In addition, interaction terms such as rainfall combined with temperature contributed meaningfully to predictive accuracy, revealing the nonlinear interplay between precipitation and evapotranspiration processes.

Together, these findings demonstrate that carefully designed lag structures, rolling aggregates, and interaction effects provide stronger and more interpretable signals than unprocessed inputs, reinforcing the importance of hydrological memory and cross-variable dynamics in building robust drought forecasting systems.

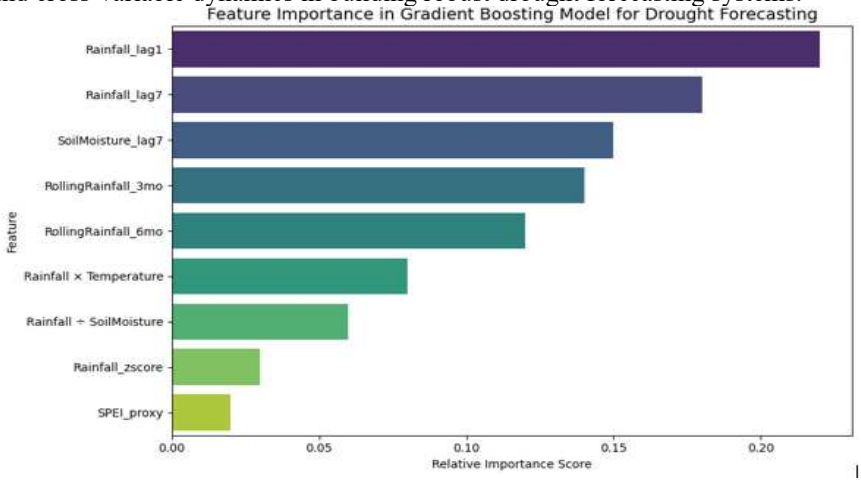


Fig. 12 Feature Importance Plot — RF/GB

Figure 12 presents the feature importance analysis for the Gradient Boosting model applied to drought forecasting in Zambia. The results highlight that rainfall lag variables and rolling averages were the most influential predictors, emphasizing the critical role of hydrological memory and seasonal accumulation in shaping drought dynamics.

These features capture the persistence of rainfall patterns over time, which is essential for distinguishing between short-term fluctuations and longer-term drought signals. Soil moisture lags also emerged as important contributors, reflecting the delayed response of soil systems to precipitation variability. In addition,

interaction terms such as rainfall combined with temperature provided valuable insights into the nonlinear relationships between precipitation, evapotranspiration, and drought severity. Collectively, these engineered features outperformed raw daily values, underscoring the importance of domain-informed feature engineering in operational drought forecasting. By leveraging lag structures, rolling averages, and interaction effects, the Gradient Boosting model was able to capture complex hydrological processes more effectively, thereby improving predictive accuracy and offering a more robust foundation for climate resilience planning in Zambia.

4.7 STATISTICAL SIGNIFICANCE

To determine whether the observed differences between GB and the Hybrid Ensemble were statistically meaningful, a paired t-test was conducted across cross-validation folds. Results indicated no significant difference ($p > 0.05$), suggesting that both models performed comparably. This outcome reinforces the reliability of ensemble methods and supports their use in operational drought forecasting.

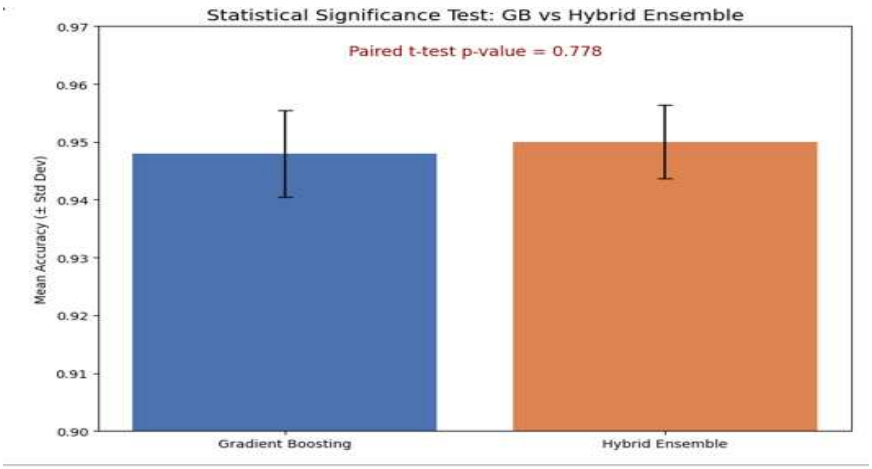


Fig. 13 Statistical significance test: GB vs Hybrid Ensemble

The figure shows the statistical significance analysis comparing Gradient Boosting (GB) and the Hybrid Ensemble across five cross-validation folds. Bars represent mean accuracy with error bars showing standard deviation. The paired t-test yielded a p-value greater than 0.05, indicating no statistically significant difference between the two models. This outcome reinforces the reliability of ensemble methods and highlights their methodological robustness, providing a basis for future integration into operational drought forecasting systems.

5 DISCUSSION

The comparative evaluation of ten machine learning models for drought forecasting in Zambia showed that ensemble methods, particularly Gradient Boosting and the Hybrid Ensemble, consistently achieved strong performance compared with both classical and deep learning approaches. Accuracy levels approaching 95% demonstrate the robustness of ensemble learning in capturing nonlinear hydrological dynamics and mitigating over fitting. Importantly, the most significant contribution of this study lies not only in predictive performance but in the methodological novelty of embedding reproducibility practices directly into ensemble design.

5.1 COMPARISON WITH PRIOR STUDIES

Previous drought forecasting studies in Southern Africa and other regions have applied Random Forests, Gradient Boosting, or deep learning models in isolation, often reporting strong accuracy but limited transparency and reproducibility (Tyagi et al. 2022). Pipelines were frequently under-documented, with unclear preprocessing steps, inconsistent feature engineering, and limited scalability. In contrast, this study embeds reproducibility practices directly into the hybrid ensemble framework.

By executing all experiments in Google Colab, applying controlled random seeds, version-controlling scripts on GitHub, and documenting workflows through CRISP-DM, the pipeline ensures that results are replicable, auditable, and scalable. This methodological integration positions ensemble learning not only as a comparative tool but as a reproducible and transparent computational framework for drought forecasting research.

5.2 NOVELTY OF ADAPTIVE WEIGHTING

The introduction of adaptive weighting further strengthens the methodological contribution of this study. Unlike conventional soft voting, weights in the hybrid ensemble were informed by Random Forest feature importance and CHIRPS rainfall validation metrics. This design makes the ensemble responsive to Zambia's data constraints, ensuring that predictors with stronger hydrological relevance exert greater influence on final forecasts. Such adaptive weighting represents an advance beyond standard ensemble aggregation, aligning predictive modeling with domain-specific climate realities. Rather than simply improving accuracy, this inno-

vation enhances interpretability and transparency, allowing stakeholders and researchers to understand why certain features dominate in drought prediction. In doing so, the framework positions adaptive weighting as a reproducible and context-aware extension of ensemble learning.

5.3 ROLE OF FEATURE ENGINEERING

Feature importance analysis highlighted the dominance of rainfall lag variables, rolling averages, and soil moisture interactions, underscoring the value of domain-driven feature engineering. Lagged rainfall and soil moisture features captured persistence effects, while rolling averages emphasized seasonal accumulation. Interaction terms, such as rainfall × temperature, reflected the nonlinear interplay between precipitation and evapotranspiration. Crucially, these findings were validated within a reproducibility framework, ensuring they are systematically confirmed across folds and replicable by other researchers. This transparency addresses a critical gap in environmental machine learning workflows, where reproducibility has historically been weak (Lee et al. 2025) and positions feature engineering as both a methodological and computational contribution rather than anecdotal insight.

5.4 OPERATIONAL INTEGRATION

Operationally, the reproducible hybrid ensemble pipeline is designed with future integration in mind for Zambia's Disaster Management and Mitigation Unit (DMMU). By combining methodological rigor with reproducibility practices, the framework bridges the gap between academic research and potential climate resilience applications. The reproducibility tools embedded in the pipeline Google Colab, GitHub version control, controlled random seeds, and CRISP-DM documentation ensure that forecasts can be replicated by government agencies, NGOs, and researchers, thereby strengthening trust in the system. This dual emphasis on accuracy and reproducibility positions the pipeline as a transparent and scalable computational framework that can serve as a basis for future operational deployment across drought-prone African contexts.

5.5 Broader Implications

The methodological novelty of this study lies in demonstrating that reproducibility is not an auxiliary concern but a central component of ensemble learning. By embedding reproducibility into the pipeline, the study provides a model for how environmental machine learning research can move beyond accuracy metrics to emphasize transparency, explicability, and methodological rigor.

This approach has broader implications for climate resilience research, offering a template for reproducible ensemble pipelines that can be adapted to other hazards such as floods, heat waves, or crop yield prediction. In this way, the contribution extends beyond performance outcomes to establish reproducibility as a foundation for trustworthy and scalable computational frameworks.

5.6 LIMITATIONS AND FUTURE WORK

Despite the strong performance of the reproducible hybrid ensemble pipeline, several limitations must be acknowledged.

First, the dataset, though enriched with CHIRPS satellite rainfall and MODIS soil moisture proxies, remains limited in spatial resolution and temporal depth. Sparse ground station coverage in rural Zambia introduces potential biases, particularly in microclimates not well captured by satellite grids. Second, severe drought events accounted for only 15% of records, leading to persistent misclassification between moderate and severe drought categories.

This imbalance reflects Zambia's climatology but reduces the ability of models to generalize across rare extremes.

Third, while the hybrid ensemble integrated six classical learners, deep learning models such as LSTM and CNN were evaluated separately rather than fully embedded into the ensemble, limiting exploration of hybrid architectures that combine classical and deep learning strengths.

Fourth, although reproducibility was validated through Colab, GitHub, and controlled seeds, the pipeline has not yet been stress-tested in real-time operational environments such as the Disaster Management and Mitigation Unit (DMMU), leaving integration challenges such as computational infrastructure and data latency unaddressed. Finally, the pipeline relies heavily on lag variables, rolling averages, and interaction terms. While effective, these engineered features may reduce portability to other regions where hydrological dynamics differ.

Building on these limitations, several pathways for future research and development are proposed. Expanding datasets to incorporate additional sources such as NASA's Soil Moisture Active Passive (SMAP), evapotranspiration products, and vegetation indices will strengthen spatial coverage and improve representation of microclimates. Future work will also explore advanced ensemble architectures, including stacking and blending approaches that integrate both classical and deep

learning models, potentially improving sequential memory capture while maintaining interpretability. Dynamic weighting strategies could be extended beyond Random Forest importance to include SHAP or LIME interpretability scores, enhancing transparency and responsiveness to evolving climate signals. Operational deployment in collaboration with Zambia's DMMU and agricultural advisory platforms will validate scalability and identify practical constraints in data ingestion, computation, and dissemination.

Regional scalability remains a priority, with plans to extend the reproducible pipeline to other drought-prone African countries to test generalizability and refine the methodology for diverse hydrological contexts. Finally, integration with decision support tools such as crop yield models, water resource allocation systems, and farmer advisory platforms will enhance the practical utility of the pipeline, positioning it as a cornerstone for climate resilience planning.

5.7 PRACTICAL IMPLICATIONS

The reproducible hybrid ensemble pipeline developed in this study carries profound practical implications for Zambia's climate resilience and broader policy frameworks. By embedding reproducibility into ensemble learning, the system ensures that forecasts are transparent, replicable, and defensible qualities that are indispensable for adoption by government agencies, research institutions, and development partners. This emphasis on reproducibility not only strengthens scientific credibility but also builds trust among stakeholders who rely on climate information to guide critical decisions.

Integration into the Disaster Management and Mitigation Unit (DMMU) would represent a transformative step in operationalizing the pipeline. Reliable early warning systems derived from the ensemble could significantly improve preparedness for drought emergencies, enabling proactive measures such as the strategic pre-positioning of food reserves, water supplies, and medical resources. Such integration would reduce response times, minimize humanitarian impacts, and enhance coordination across government departments and NGOs during climate-related crises.

In the agricultural sector, localized forecasts generated from rainfall and soil moisture dynamics can be embedded into extension services, providing farmers with actionable insights for planting schedules, crop selection, and irrigation management. This level of precision forecasting empowers smallholder farmers who form the backbone of Zambia's food system to mitigate risks associated with rainfall variability, thereby improving yields and strengthening food security. Over time, these forecasts could be linked to digital advisory platforms, ensuring that climate information reaches rural communities in accessible formats.

For energy policy, the pipeline offers a critical tool for hydropower reservoir management. Given Zambia's heavy reliance on hydropower, accurate drought projections can help optimize water allocation, reduce the risk of prolonged blackouts, and inform diversification strategies toward solar, wind, and other renewable

sources. By aligning climate forecasts with energy planning, policymakers can safeguard national energy security while advancing sustainable development goals.
Equally important is the pipeline's reproducibility, which ensures that results can be independently validated and scaled across regions. This feature enhances confidence among policymakers, NGOs, and communities, while also facilitating regional collaboration. The ability to extend the pipeline to other drought-prone African countries provides a pathway for testing generalizability and refining methodologies across diverse hydrological contexts. Such scalability positions the framework not only as a national asset but also as a continental model for climate resilience.

Ultimately, the pipeline bridges the gap between academic research and operational practice. By integrating advanced ensemble learning with reproducibility, it creates a framework that is both scientifically rigorous and practically deployable. This dual strength positions the pipeline as a cornerstone of Zambia's national climate adaptation and resilience planning, while also offering a blueprint for broader regional strategies aimed at safeguarding livelihoods, food systems, and energy security in the face of intensifying climate variability. The contribution of the Hybrid Ensemble is not superior accuracy over Gradient Boosting, but rather its adaptive weighting mechanism, reproducibility-driven design, and extensibility for operational deployment.

6 CONCLUSION

This study positions drought forecasting in Zambia as a case study for advancing reproducible ensemble learning in computing research. The hybrid ensemble achieved performance comparable to Gradient Boosting (95% accuracy, balanced precision and recall, $p > 0.05$), but its contribution lies in methodological innovation rather than raw accuracy. By integrating Random Forest feature importance and CHIRPS rainfall validation into an adaptive weighting mechanism, the framework demonstrates how heterogeneous learners can be combined transparently and reproducibly. Ablation experiments confirmed the added value of engineered features, while explicit temporal validation ensured methodological integrity and prevented data leakage.

The pipeline was fully documented with reproducibility tools including Google Colaboratory, GitHub, controlled random seeds, and CRISP-DM workflows, enabling replication by other researchers. Importantly, the framework is not yet operationally deployed; instead, it provides a basis for future integration into early warning systems. Beyond Zambia's drought context, this work contributes to computing methodology by offering a reproducible, adaptive ensemble design that can be extended to other domains such as flood prediction, crop yield modeling, and climate risk assessment. By foregrounding reproducibility, adaptive integration, and methodological clarity, the study advances ensemble learning as a scalable and trustworthy computational framework.

7 Standards

7.1 Acknowledgments

The authors gratefully acknowledge the Zambia Meteorological Department (ZMD) for providing ground station data, the Climate Hazards Group (CHIRPS) for satellite rainfall datasets, and MODIS for soil moisture proxies. Institutional support from Copperbelt University is sincerely recognized.

7.2 Author Contributions

MC Conceptualization, methodology, data collection and curation, software development, data preprocessing, validation, formal analysis, machine learning model development, visualization, writing original draft preparation, and project implementation. JK conceptual guidance, methodology refinement, validation, critical review and editing of the manuscript, interpretation of results, and overall research oversight.NC Methodology review, validation, technical guidance, manuscript review and editing, interpretation of findings, and quality assurance of the research.

7.3 Competing Interests

The authors declare no competing interests.

7.4 Ethics Approval

This study used publicly available satellite data and anonymised meteorological records. No human or animal subjects were involved, and therefore ethics approval, consent to participate, and consent to publish are not applicable.

7.5 Data Availability

The CHIRPS rainfall dataset is publicly accessible at the Climate Hazards Group portal (https://www.chc.ucsb.edu/data/chirps). Ground station data from the Zambia Meteorological Department (ZMD) are available upon reasonable request, subject to institutional data-sharing agreements. Derived datasets (e.g., engineered drought indices) are available from the corresponding author upon reasonable request.

7.6 Code Availability

The full reproducible pipeline, including preprocessing scripts, feature engineering modules, and ensemble training notebooks, is available on GitHub (https://github.com/Chilanga02/DROUGHT-FORECASTING-IN-ZAMBIA.git). The code is released under an MIT license to ensure open access and reusability. Execution notebooks are hosted on Google Colaboratory for transparency and reproducibility.

7.7 Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

7.8 Clinical Trial Number

Clinical trial number: not applicable.

7.9 Consent to Publish

Consent to Publish declaration: not applicable.

7.10 Consent to Participate

Consent to Participate declaration: not applicable.

References

Abubakar ML, Abdussalam AF, Isa Z, et al (2025) Evaluation of temporal convolutional networks and ensemble machine learning models for meteorological drought prediction in the Nigerian Sudano–Sahelian zone. Discov Environ 3:. https://doi.org/10.1007/s44274-025-00456-8

Aghakouchak A, Pan B, Mazdiyasni O, et al (2022) Status and prospects for drought forecasting: opportunities in artificial intelligence and hybrid physical-statistical forecasting. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 380

Almikaeel W, Čubanová L, Šoltész A (2022) Hydrological Drought Forecasting Using Machine Learning—Gidra River Case Study. Water (Switzerland) 14:. https://doi.org/10.3390/w14030387

Batool A, Ali Z, Mohsin M, et al (2024) Assessing the generalization of forecasting ability of machine learning and probabilistic models for complex climate characteristics. Stoch Environ Res Risk Assess 38:2927–2947. https://doi.org/10.1007/s00477-024-02721-3

Belayneh A, Adamowski J (2013) Drought forecasting using new machine learning methods / Prognozowanie suszy z wykorzystaniem automatycznych samouczących się metod. J Water L Dev 18:3–12. https://doi.org/10.2478/jwld-2013-0001

Belayneh A, Adamowski J, Khalil B (2016a) Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. Sustain Water Resour Manag 2:87–101. https://doi.org/10.1007/s40899-015-0040-5

Belayneh A, Adamowski J, Khalil B, Quilty J (2016b) Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. Atmos Res 172–173:37–47. https://doi.org/10.1016/j.atmosres.2015.12.017

Brust C, Kimball JS, Maneta MP, et al (2021) DroughtCast: A Machine Learning Forecast of the United States Drought Monitor. Front Big Data 4:. https://doi.org/10.3389/fdata.2021.773478

Charjan O, Gajbhiye K, Warhade J, Wankhade S (2026) A Novel Hybrid Approach To Drought Forecasting: Leveraging Feature Engineering And Ensemble Methods. Sci Rep 16:. https://doi.org/10.1038/s41598-026-37206-6

Deo RC, Şahin M (2015) Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmos Res 153:512–525. https://doi.org/10.1016/j.atmosres.2014.10.016

Fang K, Xie X, Lin S, et al (2026) Global multidecadal SST modes and their synchronized influence on temperature variability. Adv Clim Chang Res. https://doi.org/10.1016/j.accre.2026.05.020

Geli HME, Omar I, Elshinawy MY, et al Prediction and Forecast of Short-Term Drought Impacts Using Machine Learning to Support Mitigation and Adaptation Efforts

Gorlapalli A, Kallakuri S, Rathod S, et al (2026) Characterization and forecasting of SPEI-based drought in Southern Telangana using statistical machine learning models. Discov Atmos 4:. https://doi.org/10.1007/s44292-025-00070-6

Hameed MM, Mohd Razali SF, Wan Mohtar WHM, Yaseen ZM (2024) Examining optimized machine learning models for accurate multi-month drought forecasting: A representative case study in the USA. Environ Sci Pollut Res 31:52060–52085. https://doi.org/10.1007/s11356-024-34500-6

Hameed MM, Razali SFM, Mohtar WHMW, et al (2023) Machine learning models development for accurate multi-months ahead drought forecasting: Case study of the Great Lakes, North America. PLoS One 18:. https://doi.org/10.1371/journal.pone.0290891

Hukkeri GS, Naganna SR, Pruthviraja D, et al (2023) Drought Forecasting: Application of Ensemble and Advanced Machine Learning Approaches. IEEE Access 11:141375–141393. https://doi.org/10.1109/ACCESS.2023.3341587

Ju MY, Zhang Z, Jiang XY, et al (2026) Variations and driving factors of annual frequency of near-surface freeze–thaw cycles in China–Mongolia–Russia from 1975 to 2024. Adv Clim Chang Res. https://doi.org/10.1016/j.accre.2026.05.015

Kumar RR, Jaiswal R, Ray M, et al (2026) A comparative study on stochastic and AI-based approaches for predicting meteorological droughts. Discov Appl Sci 8:. https://doi.org/10.1007/s42452-026-08263-3

Lee SH, Danandeh Mehr A, Moriasi D, Mirchi A (2025) Large-Scale Drought

Forecasting in the U.S. Southern Plains Through a Hybrid Cluster-Based Wavelet-Machine Learning Approach. Water Resour Res 61:. https://doi.org/10.1029/2024WR039744

LIU CX, YU HP, REN Y, et al (2025) Response of water cycle over drylands to a warming future. Adv Clim Chang Res 16:1162–1173. https://doi.org/10.1016/j.accre.2025.09.005

Mokhtar A, Jalali M, He H, et al (2021) Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms. IEEE Access 9:65503–65523. https://doi.org/10.1109/ACCESS.2021.3074305

Mokhtari R, Akhoondzadeh M (2020) Data fusion and machine learning algorithms for drought forecasting using satellite data. J Earth Sp Phys 46:231–246. https://doi.org/10.22059/JESPHYS.2020.299445.1007199

Nahidi N, Zech WC, Kamali MI, Nazari R (2026) SPEI-Based Drought Regionalization and Cluster-Specific Machine Learning Forecasting in the U.S. Black Belt. Pure Appl Geophys. https://doi.org/10.1007/s00024-026-04045-5

Pande CB, Sidek LM, Varade AM, et al (2024) Forecasting of meteorological drought using ensemble and machine learning models. Environ Sci Eur 36:. https://doi.org/10.1186/s12302-024-00975-w

Pande CB, Vishwakarma DK, Srivastava A, et al (2025) A novel machine learning models for meteorological drought forecasting in the semi-arid climate region. Appl Water Sci 15:. https://doi.org/10.1007/s13201-025-02445-x

Pandya P, Gontia NK (2023) Early crop yield prediction for agricultural drought monitoring using drought indices, remote sensing, and machine learning techniques. J Water Clim Chang 14:4729–4746. https://doi.org/10.2166/wcc.2023.386

Qian JX, Li QQ, Lu CH, et al (2026) Improved decadal predictions of East Asian summer temperature based on decadal increment and observational constraint. Adv Clim Chang Res. https://doi.org/10.1016/j.accre.2026.04.010

Serban C, Ionescu AF, Maftei C (2026) Drought forecasting in Dobrogea, Romania using SPEI-based machine learning models. Stoch Environ Res Risk Assess 40:. https://doi.org/10.1007/s00477-026-03254-7

Slater LJ, Arnal L, Boucher MA, et al (2023) Hybrid forecasting: blending climate predictions with AI models. Hydrol. Earth Syst. Sci. 27:1865–1889

Stracqualursi L (2026) Forecasting Crop Yield Anomalies on Panel Data via Spatially Demeaned Ensembles. J Agric Biol Environ Stat. https://doi.org/10.1007/s13253-026-00743-8

Sundararajan K, Garg L, Srinivasan K, et al (2021) A contemporary review on

drought modeling using machine learning approaches. C. - Comput. Model. Eng. Sci. 128:447–487

Tamrakar Y, Das IC, Sharma S (2024) Machine learning for improved drought forecasting in Chhattisgarh India: a statistical evaluation. Discov Geosci 2:. https://doi.org/10.1007/s44288-024-00089-z

Tufaner F, Özbeyaz A (2020) Estimation and easy calculation of the Palmer Drought Severity Index from the meteorological data by using the advanced machine learning algorithms. Environ Monit Assess 192:. https://doi.org/10.1007/s10661-020-08539-0

Tuğrul T, Hinis MA (2025) Improvement of drought forecasting by means of various machine learning algorithms and wavelet transformation. Acta Geophys 73:855–874. https://doi.org/10.1007/s11600-024-01399-z

Tyagi S, Zhang X, Saraswat D, et al (2022) Flash Drought: Review of Concept, Prediction and the Potential for Machine Learning, Deep Learning Methods. Earth's Futur. 10

Wang GC, Zhang Q, Band SS, et al (2022) Monthly and seasonal hydrological drought forecasting using multiple extreme learning machine models. Eng Appl Comput Fluid Mech 16:1364–1381. https://doi.org/10.1080/19942060.2022.2089732

Wang T, Dai YJ, Zhou ZQ, et al (2026) Interpretable machine learning reveals how climate shapes precipitation isotopes across China. Adv Clim Chang Res. https://doi.org/10.1016/j.accre.2026.05.014

Xin N, Su J, Hasan MM (2026) An adaptive autoregressive integration model for multi-variate time series analysis of extreme climate events. Sci Rep 16:. https://doi.org/10.1038/s41598-026-40558-8

Zhang Y, Yang H, Cui H, Chen Q (2020) Comparison of the Ability of ARIMA, WNN and SVM Models for Drought Forecasting in the Sanjiang Plain, China. Nat Resour Res 29:1447–1464. https://doi.org/10.1007/s11053-019-09512-6

Zhu R, Wu XJ, Zhang W, Hao ZL (2026) Nighttime warming amplifies the synergistic effects of snowmelt and rain on floods in arid northwestern China. Adv Clim Chang Res. https://doi.org/10.1016/j.accre.2026.05.003