RESEARCH **Open Access**

Climate-driven maize yield prediction across Uganda's ZARDI zones using pooled learning and uncertainty-aware modeling

Danison Taremwa[1,5*], Emmanuel Ahishakiye[2], Aggrey Obbo[3], Paul Kategaya Kisozi[4] and Fred Kaggwa[5]

*Correspondence:
Danison Taremwa
taremwa.danison@gmail.com
[1]Department of Computer Science, Kyambogo University, Kampala, Uganda
[2]Department of Networks, Data Science and Artificial Intelligence, Kyambogo University, Kampala, Uganda
[3]Department of Software Engineering, Mbarara University of Science & Technology, Mbarara, Uganda
[4]Department of Environmental Science, Kyambogo University, Kampala, Uganda
[5]Department of Computer Science, Mbarara University of Science & Technology, Mbarara, Uganda

Abstract

Accurate maize yield prediction is essential for agricultural planning and food security, yet remains challenging in data-scarce, heterogeneous agroecological settings. This study investigates climate-driven maize yield prediction across Uganda's Zonal Agricultural Research and Development Institute (ZARDI) regions using a pooled modeling framework that mitigates the problem of extremely small zone-specific sample sizes. Seasonal climatic variables, including rainfall, temperature, solar radiation, and soil moisture, collected from 2018 to 2021, were integrated with zone indicators and evaluated under a strict leave-one-year-out (LOYO) cross-validation protocol to prevent temporal information leakage. Regularized linear regression (Ridge) achieved the most reliable generalization performance, with a root mean squared error (RMSE) of approximately 0.98 t ha^{-1} and mean absolute error (MAE) of 0.57 t ha^{-1}, marginally outperforming global-mean (RMSE ≈ 0.99 t ha^{-1}) and ZARDI-mean (RMSE ≈ 1.05 t ha^{-1}) baselines. More complex models, including a constrained XGBoost regressor, did not yield consistent performance gains under LOYO evaluation. Zone-specific analysis revealed substantial heterogeneity, with some regions exhibiting moderate predictive skill (MAE < 0.25 t ha^{-1}), while others showed weak climate–yield coupling and negative coefficients of determination. To address overconfidence in the face of data scarcity, prediction uncertainty was explicitly quantified using conformal prediction intervals based on out-of-fold residuals. Global prediction intervals achieved empirical coverage of 91.7% and 96.3% at the 90% and 95% nominal levels, respectively, indicating well-calibrated uncertainty estimates. Zone-specific interval widths varied markedly, ranging from approximately 0.84 t ha^{-1} in better-performing regions to over 4.6 t ha^{-1} in the most uncertain zones. Model interpretability analysis using SHapley Additive exPlanations identified rainfall, soil moisture, solar radiation, and seasonal timing as the dominant drivers of yield predictions, while zone identifiers contributed minimally. The proposed framework provides a transparent, defensible basis for yield assessment and underscores the need to integrate additional agronomic and remotely sensed data to improve predictive skill in regions with weak climate–yield relationships.

Highlights

- Pooled cross-zone modelling predicts maize yield across Uganda's ZARDI zones despite very small zone-level samples.
- Leave-one-year-out validation shows pooled Ridge regression performs best (RMSE ≈ 0.98 t ha^{-1}; MAE ≈ 0.57 t ha^{-1}) and slightly outperforms mean baselines.
- Conformal prediction yields calibrated uncertainty (≈ 91.7% and ≈ 96.3% coverage at the 90% and 95% nominal levels, respectively) with clear zone-to-zone variation.
- SHAP interpretability identifies rainfall, soil moisture, solar radiation, and seasonal timing as dominant drivers, while zone identifiers contribute minimally.

Keywords Maize yield prediction, Data-efficient machine learning, Explainable AI, Climate–yield interactions, Agro-ecological zoning

1 Introduction

Global population growth poses a significant threat to food security. The United Nations projects that the world's population will reach about 9.8 billion by 2050, roughly 30% higher than today [1, 2]. Much of the projected growth is expected in developing regions, particularly sub-Saharan Africa (SSA), where the annual population growth rate exceeds 2%, compared to the global rate of 0.85%, even as these countries have limited adaptive capacity [3]. Such demographic shifts will heighten food demand, requiring a 60–70% increase in food production by 2050 to meet the needs of the projected larger population [4, 5]. However, global crop yields are stagnating, with declines of 24–39% across areas cultivated with maize, rice, wheat, and soybeans, partly due to climate variability, including droughts, heatwaves, and floods [6]. For instance, a 1 °C increase in nighttime temperatures can reduce maize yields by 10–15% [7]. Climate models further project continued warming and changes in rainfall patterns in East Africa, leading to significant declines in staple crops, such as maize, with Uganda forecasting yield reductions of 7–10% over the next decade under unfavorable climate conditions [8, 9]. Broader trends suggest that, without intervention, crop productivity in SSA will reduce by 50% by 2030, driven by climate unpredictability, underscoring the need for reliable yield forecasts to support food security planning [10].

Within this context, maize is crucial to food security in Uganda, serving as both a staple food and a significant cash crop, accounting for approximately 1.7% of the national export value [11]. It is grown by at least 86% of smallholder farmers and covers about 23% of the land across all ZARDI zones due to its adaptability to different climatic conditions [12]. Despite its importance, national maize yield remains low and highly variable, averaging 2.3–2.7 t/ha, with substantial zonal disparities that are well below the global average of 5–6 t/ha [13, 14]. Addressing yield stagnation and variability requires timely, accurate, and spatially explicit in-season yield information to guide farmers and policymakers in making planting decisions, optimizing input use, and developing early-warning systems [15]. However, predicting maize yield remains highly complex due to the interactions among numerous heterogeneous, nonlinear factors, including environmental variability, crop genetics, pest pressure, and crop health status [16]. These interactions vary substantially across space and time, making yield prediction a fundamentally data-intensive and modeling-challenging problem [17].

1.1 Motivation

Historically, maize yield prediction (MYP) in Uganda relied heavily on field-based methods, including farmer surveys, manual sampling, and empirical agronomic models [18]. These current approaches rely on basic statistical models conducted by extension agents through farmer interviews and crop-cut sampling, which often result in slow, labor-intensive, and inaccurate results. For instance, [19] conducted a study across Eastern Uganda, in which they regressed maize yield on Sentinel-2 satellite-derived vegetation indices calibrated with in-field measurements from crop cuts. This approach explained approximately 54% of the variance in yield ($R^2 = 0.54$). Motivated by these limitations, process-based crop simulation models, including FAO's AquaCrop and DSSAT's CERES-Maize, have improved the accuracy of maize yield predictions by modeling interactions between environmental factors and yield. These models simulate maize growth based on weather, soil, cultivar, and management inputs, providing insights through "what-if" analyses that traditional statistics cannot offer. For instance, [20] used the DSSAT-CERES-Maize model to predict a 9.6–11.8% decrease in national maize yields by 2030 and a 13.3–29.4% decrease by 2090 due to rising temperatures and moisture deficits. However, the broader application of CSMs is costly and requires specialized knowledge and intensive efforts to collect daily meteorological data across many individual farms in developing countries [21].

The increasing availability of satellite observations and advances in computing hardware have fostered the widespread adoption of data-driven machine learning (ML) approaches for maize yield prediction (MYP) [22]. These models can capture and learn complex, nonlinear climate-yield relationships from multimodal inputs, but their reliability depends on both sufficiently extensive and diverse datasets and the evaluation design [23, 24]. In data-scarce settings, highly flexible ML models can overfit and may generalize poorly across years, whereas regularized models, such as Ridge regression, strike a more robust balance between flexibility and stability [25]. These considerations are especially important for decision support, where point accuracy alone is insufficient. Models should generalize well to unseen seasons and provide calibrated uncertainty estimates alongside interpretable evidence of dominant climatic drivers, including their relative importance and influence direction, particularly in climate-sensitive, rainfed production systems [26, 27]. The evaluation strategy must also align with the forecasting objective, because random train-test splits can mix years and introduce temporal leakage, thereby inflating performance. Inter-annual forecasting is therefore more appropriately assessed using strict time-based validation, such as leave-one-year-out (LOYO) cross-validation [25].

In Uganda, agroecological diversity across ZARDI regions complicates yield prediction because current research often relies on national datasets, masking zone-specific differences crucial for tailored decision support [9]. Similarly, most Uganda-focused studies still emphasize point prediction in non-temporal evaluation designs and rarely integrate uncertainty quantification and explainability at the zone level. Consequently, evidence that jointly (i) tests true inter-annual forecasting using leakage-free temporal validation (e.g., LOYO), (ii) reports calibrated predictive uncertainty (prediction intervals), and (iii) delivers stable, zone-resolved interpretations of dominant climatic drivers, including both relative importance and effect direction, remains limited. This limitation is particularly acute because zone-level annual yield records are typically short (2018–2021), increasing the risk of overfitting and the uncertainty in inferred driver effects. This evidence gap constrains ZARDI-targeted planning and seasonal advisory decisions, which require both reliable predictions and transparent climate-based justification.

To address these gaps, this study develops an explainable framework for maize yield prediction that integrates pooled ML modeling, rigorous temporal validation, uncertainty quantification, and ZARDI-specific interpretation of climatic drivers for the period from 2018 to 2021. Specifically, the objectives include: (i) developing pooled yield prediction models that learn climate-yield relationships under limited data and evaluating them using a leave-one-year-out (LOYO) protocol; (ii) quantifying predictive uncertainty using calibrated conformal prediction intervals under the same LOYO evaluation; and (iii) ensuring consistent attribution of significant climatic drivers using SHAP-based explanations to elucidate feature importance and the direction of influence across zones.

1.2 Contributions

This study makes the following key contributions to climate-driven crop yield modeling in data-scarce, heterogeneous agro-ecological settings:

a) Pooled, Regularized Yield Modeling under Severe Data Scarcity: This study proposes a pooled maize yield prediction framework across Uganda's ZARDI regions that explicitly addresses extremely small zone-specific sample sizes. By employing regularized regression and strict leave-one-year-out validation, the approach avoids overfitting and provides a statistically defensible alternative to zone-specific machine learning models trained on limited observations.
b) Rigorous Benchmarking with Strong Baselines and Temporal Validation: The proposed framework is systematically benchmarked against global-mean and ZARDI-mean predictors, as well as more complex machine learning models, under leakage-free temporal validation. The results quantitatively demonstrate that simpler, regularized models can generalize more reliably than flexible ensemble methods in data-constrained agricultural settings.
c) Explicit Zone-Aware Uncertainty Quantification for Yield Prediction: A key contribution is the integration of conformal prediction intervals to quantify both global and zone-specific uncertainty in maize yield estimates. The framework achieves near-nominal empirical coverage under strict temporal hold-out and reveals substantial heterogeneity in predictive uncertainty across ZARDIs, enabling uncertainty-aware interpretation and decision-making.
d) Stable and Interpretable Climate Driver Analysis: The study provides a cautious and robust interpretability analysis using SHapley Additive exPlanations applied to a pooled, regularized model. The analysis identifies rainfall, soil moisture, solar radiation, and seasonal timing as the dominant yield drivers, while demonstrating that zone identifiers contribute minimally, confirming that predictions are climate-driven rather than memorized zone-based.

1.3 Paper outline

The remainder of this paper is organized as follows. Section 2 reviews related work on climate-driven crop yield prediction, with particular emphasis on machine learning approaches, validation strategies, and uncertainty quantification in data-scarce agricultural settings. Section 3 describes the study area, datasets, and methodology, including data preprocessing, pooled modeling with regularized regression, baseline benchmarking, and the leave-one-year-out evaluation protocol. Section 4 presents the experimental results, covering overall model performance, comparison with baseline predictors, zone-specific performance across ZARDIs, uncertainty quantification using conformal prediction intervals, and model interpretability through SHapley Additive exPlanations. Section 5 discusses the results in relation to existing literature, highlighting methodological implications, limitations, and practical relevance. Finally, Sect. 6 concludes the paper by summarizing the main findings, outlining key limitations, and identifying directions for future research. The upcoming section discusses materials and methods [28].

2 Related work

2.1 Climate-based crop yield forecasting

Crop yields, and maize in particular, are highly sensitive to interannual climate variability and extremes. In rainfed systems, variations in rainfall (e.g., delayed onset, intra-season droughts) and heat stress during flowering commonly limit maize development and grain filling [2]. In SSA, rainfall effects vary; for example, in regions such as Benue State, Nigeria, increased rainfall can enhance yields when water is limited [29], while excess rainfall can adversely affect yields in areas such as KwaZulu-Natal, South Africa [30]. In Uganda, maize productivity thrives under moderate temperatures (≈ 20–27 °C) in the central, western, and eastern highlands but declines in the north, where temperatures vary more widely (≈18–35 °C) [9]. Projections for East Africa anticipate temperature increases of 1.7–5.4 °C and a 5–20% reduction in annual rainfall by the century's end. In the short term, Uganda is expected to face maize yield losses of 7–10% over the next decade [8]. These challenges underscore the need for localized, evidence-based forecasting that links climate variability to agricultural outcomes, and ML has emerged as a viable method for leveraging climate and satellite-derived data [2].

2.2 Machine learning in small-sample yield prediction

Machine learning methods have become central to crop yield prediction because they can learn complex nonlinear relationships from climate time series, remote sensing data, and other sources [24, 31]. In rich-data settings, tree-based ensembles and deep learning architectures (e.g., convolutional or recurrent neural networks) can effectively exploit high-dimensional satellite and meteorological inputs to improve yield forecasts [32]. For example, [33] employed an ensemble ML model to predict maize yield in India, using a climatic dataset spanning 2016–2018. Results indicated improved performance for ensemble methods, with coefficient of determination (R^2) values for XGBoost (0.8509), random forest (RF) (0.7952), and Voting (0.7599), compared to single models like Decision Tree (0.4516) and Linear (0.3470). Nevertheless, ensemble methods can falter when data are extremely limited. With very few seasonal observations, boosting models may "memorize" idiosyncratic year-to-year patterns rather than learning true climate–yield dynamics [34]. In such cases, ensemble complexity can amplify overfitting, especially if important covariates (management, pests, soil fertility) are missing [35].

However, ML models do not automatically generalize well under severe data scarcity [25]. When sample sizes are small, or predictors are incomplete (e.g., missing management or soil covariates), highly flexible models may overfit to noise and produce unstable out-of-sample forecasts [31]. Accordingly, multiple studies have observed that simpler, regularized regression models can outperform complex learners on limited national or zone-level datasets [36]. In practice, robust forecast pipelines often rely on ridge regression or lasso to stabilize estimates in the face of collinearity and weak signals. These approaches preserve interpretability and avoid the overly optimistic in-sample fits that can arise with unregularized ML on scarce climate–yield data [35].

2.3 Temporal validation and generalization

Proper evaluation is critical in yield forecasting research because standard random splits can mix data from different years, leading to temporal leakage and inflated performance [37]. Since annual climate and yield data are temporally autocorrelated, true interannual forecasting requires time-aware validation. LOYO cross-validation, in which data from a held-out year are excluded from training, is widely recommended for assessing generalization to future seasons [38]. By withholding entire years, LOYO prevents a model from exploiting shared year-level effects or trend information when predicting unseen seasons. For example, [25] demonstrates that rigorous LOYO evaluation is needed to obtain realistic accuracy estimates in regional forecasts. Time-based validation is especially important in small-zone datasets, where the choice of held-out year can strongly affect results; failure to apply LOYO often yields overly optimistic estimates of skill. A number of crop forecasting studies now explicitly note that temporally strict validation is essential for credible claims [25, 38].

2.4 Uncertainty quantification in yield forecasting

Crop yield prediction is inherently uncertain due to measurement noise, missing inputs (e.g., management, pests, soil), and spatial heterogeneity [25]. Despite this, many studies report only point forecasts. Increasingly, decision-support applications demand calibrated uncertainty estimates (e.g., prediction intervals) to convey forecast reliability [38]. A promising framework for uncertainty quantification is conformal prediction, a non-parametric, model-agnostic technique. Conformal prediction has emerged as a practical, distribution-free approach to generate such intervals with finite-sample guarantees [27]. Conformal methods require no specific error distribution and can be applied post-hoc to any regression model. In agriculture, conformal intervals have demonstrated well-calibrated coverage: empirical studies report that conformal prediction intervals for yield estimates meet nominal confidence levels with minimal assumptions [38]. This makes them especially attractive for data-scarce settings such as the ZARDI zones, where users need an explicit signal indicating when model predictions may be unreliable. Incorporating calibrated intervals helps ensure that yield forecasts in risk-averse decisions are accompanied by defensible uncertainty bounds [38].

2.5 Explainable AI for climate–yield interpretation

Model interpretability is increasingly recognized as crucial in agricultural ML, where stakeholders require insight into the drivers behind predictions. Explainable AI (XAI) techniques link model outputs to input features, enabling users to understand which climate variables influence yields and whether these align with agronomic knowledge [26]. Shapley-value methods, notably SHAP, have become popular for global and local explanations of yield models. SHAP assigns additive importance scores to each feature, reflecting its contribution to predictions. Alternative methods, such as LIME, similarly provide local, model-agnostic explanations [39]. But it was excluded due to its instability and inconsistent global interpretability in small, noisy datasets. Recent crop-yield studies have applied SHAP to ensemble and deep models, revealing that precipitation and temperature typically dominate yield forecasts [40]. Importantly, the reliability of these explanations depends on the quality of the model and the data. In small-sample regimes, explanations from highly flexible models can be unstable or spurious. Hence, current practice advocates pairing XAI with robust modeling and rigorous validation to ensure attribution analyses reflect true climate–yield relationships rather than artifacts. In the context of this work, explainability serves to verify that the pooled model's predictions are driven by climate covariates (e.g., rainfall, soil moisture) rather than by overfitting to zone identifiers or temporal trends [26].

In conclusion, the literature indicates that while advanced ML and ensemble methods can yield strong performance in data-rich yield prediction settings, robust forecasting in heterogeneous, data-scarce systems requires careful alignment between model capacity, evaluation strategy, and interpretability approach. Key insights motivating this study are: temporally strict validation is essential to avoid leakage and inflated accuracy [37]; regularized models can generalize more reliably than high-capacity ensembles under limited observations; uncertainty quantification is necessary for defensible decision support [38]; and stable explainability methods such as SHAP should be applied within a rigorous modeling framework to provide credible climate–yield interpretations [26]. These insights directly motivate the pooled learning strategy, the LOYO evaluation, the conformal uncertainty intervals, and the SHAP-based interpretability adopted in the proposed ZARDI-wise framework. The upcoming section discusses materials and methods.

3 Materials and methods

3.1 The study area and spatial distribution of ZARDI zones in Uganda

The study was conducted across all ten ZARDI zones in Uganda. Each ZARDI, identified by its acronym, serves a specific region with multiple districts. This zonal distribution ensures broad geographic representation and captures Uganda's diverse agro-ecological zones. Uganda's tropical climate, situated between latitudes 1° South and 4° North of the equator, makes it an ideal location for agriculture. The ZARDIs are strategically located in the country's diverse agro-ecological zones, as shown in Fig. 1, providing tailored agricultural research and support across various regions, as illustrated in the map below [41].

3.2 Dataset description

This study employed an integrated climate–yield dataset compiled for Uganda's Zonal Agricultural Research and Development Institute (ZARDI) regions to support climate-driven maize yield prediction under data-scarce conditions. Each observation corresponds to a single maize growing season for a given ZARDI and comprises one yield outcome paired with a set of climatic, temporal, and structural predictors. The target variable is maize yield, expressed in tonnes per hectare (t ha^{-1}) and representing seasonal average yields aggregated at the ZARDI level. Across the dataset, maize yields exhibit moderate inter-annual variability and systematic differences among zones, reflecting underlying agro-ecological heterogeneity and differences in production environments. Predictor variables include seasonal climatic indicators selected based on

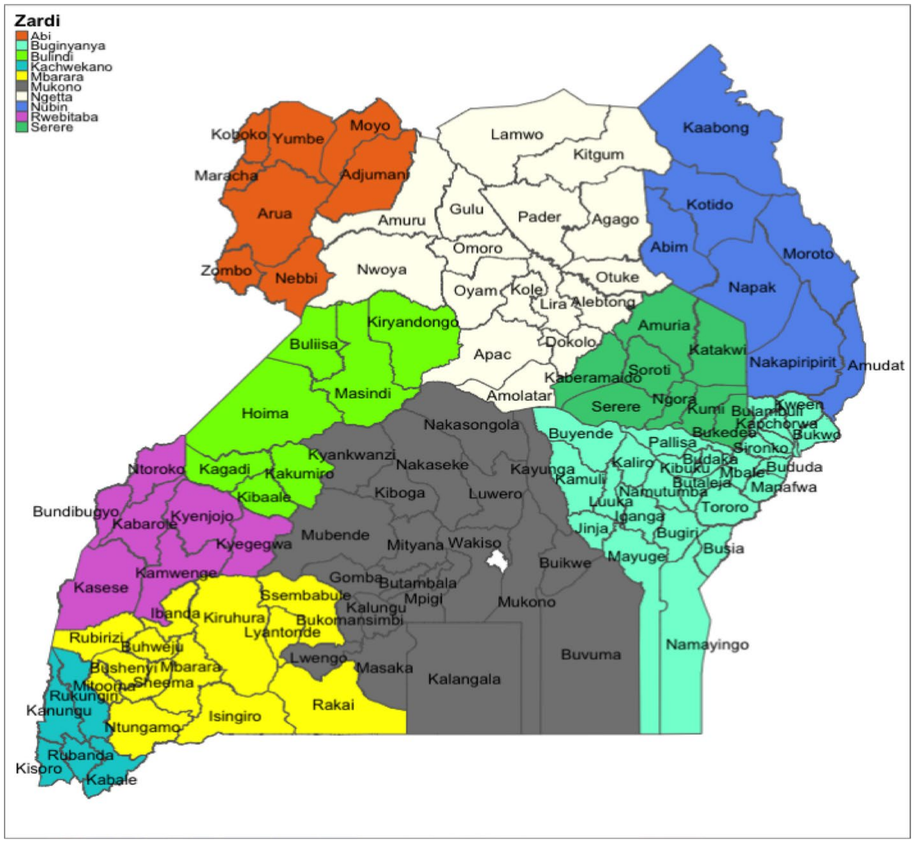

Fig. 1 The study area

their agronomic relevance to maize growth and productivity. These comprise total seasonal rainfall (mm), mean soil moisture, mean solar radiation, maximum temperature (°C), and minimum temperature (°C), all aggregated over the cropping period to capture conditions relevant to crop development. To enhance interpretability while maintaining model simplicity, derived features, such as the temperature range (the difference between the maximum and minimum temperatures) and the rainfall–temperature ratio, were constructed as proxies for thermal stress and moisture availability. Temporal indicators (year and season) were included to account for seasonal and long-term trends, while ZARDI identifiers were encoded as categorical variables to represent structural differences between zones and enable pooled modeling across regions. The dataset is characterized by a limited number of observations per ZARDI, reflecting the short temporal span of available historical yield records. On average, each zone contains only a few seasonal observations across consecutive years, rendering zone-specific model training statistically unstable. To mitigate this constraint, a pooled analytical framework was adopted to leverage shared information across zones while preserving spatial heterogeneity through zone indicators. Prior to modeling, the dataset underwent systematic quality control and preprocessing.

3.2.1 Maize yield data

Maize yield data, expressed in tons per hectare (t/ha), were obtained from the Uganda Bureau of Statistics through the 50 × 2030 Initiative, a global agricultural data program led by the Food and Agriculture Organization and the World Bank. The data originate from the Annual Agricultural Survey, which applies standardized methodologies, including structured farmer interviews and crop-cut measurements, to ensure national representativeness and internal consistency across regions and years. Each ZARDI contributed eight yield observations, corresponding to Season 1 and Season 2 for each of the four study years. Yield levels varied substantially across zones, reflecting Uganda's agro-ecological heterogeneity. Mean seasonal yields ranged from approximately 0.6 t/ha in drought-prone regions such as Nabuin to above 2.0 t/ha in relatively stable regions such as Buginyanya and Mbarara, with notable inter-seasonal variability in zones affected by erratic rainfall and temperature extremes.

3.2.2 Climatic and environmental variables

Agro-climatic variables were sourced for the same period (2018–2021) from NASA POWER and Google Earth Engine, ensuring temporal alignment with the maize yield data. The variables included soil moisture, rainfall, solar radiation, maximum temperature, and minimum temperature, all of which are well-established determinants of maize growth and productivity. Climatic variables were aggregated to the ZARDI season level to match the temporal resolution of the yield observations and to reduce noise from short-term weather fluctuations.

3.2.3 Data quality control and missing data handling

Several quality-control procedures were implemented to ensure data reliability and reproducibility. All datasets were first screened for missing values, inconsistencies, and duplicate records. Missing climatic observations (less than 5% of total records) were imputed using median values computed within the same ZARDI and season, thereby preserving regional climatic characteristics. Yield records outside agronomically plausible ranges were reviewed and removed following domain-expert validation. Duplicate entries were removed, and all continuous predictors were standardized to a mean of 0 and unit variance to facilitate stable estimation in regularized regression models and reduce sensitivity to differences in predictor scales. Exploratory inspection confirmed the absence of extreme outliers after preprocessing, although moderate correlations among temperature-related variables motivated the use of regularization in subsequent modeling. The resulting dataset provides a unified, zone-aware representation of maize yield and seasonal climatic conditions, suitable for pooled learning, conservative temporal validation, uncertainty quantification, and interpretable modeling in data-scarce agricultural contexts. Descriptive statistics summarizing the distributions of all variables, including mean, standard deviation, quartiles, and range, are presented in Table 1, providing transparency into the data's scale and variability.

Table 1 Presents descriptive statistics of the dataset

Features	Count	Mean	Std	Min	25%	50%	75%	Max
SOIL_MOISTURE	80.0	0.618821	0.152066	0.407500	0.488750	0.600833	0.732708	0.9475
RAINFALL	80.0	135.106479	54.685118	23.730000	97.448333	131.833750	156.665625	305.8600
SOLAR_RAD	80.0	134.713458	2.828056	128.476667	132.944167	134.728750	136.723750	145.5375
MAX_TEMP	80.0	29.598656	3.263544	23.790000	26.850000	28.782500	32.451875	35.9825
MIN_TEMP	80.0	15.951246	2.099046	11.182500	14.345000	16.422500	17.506875	19.4525
YIELD	80.0	1.790000	0.973523	0.600000	1.300000	1.600000	2.100000	8.7000

3.2.4 Dataset characteristics and constraints

The final dataset comprises 80 season-level observations, evenly distributed across ZARDI zones but with limited temporal depth (Table 1). While this balanced structure enables fair zone-wise comparisons, the small number of observations per ZARDI limits statistical power and poses challenges for model generalization, particularly in zones with high inter-seasonal variability, such as Nabuin and Rwebitaba. These constraints reflect the current realities of nationally harmonized agricultural data availability in Uganda and are explicitly considered in the modeling strategy, evaluation protocol, and interpretation of results.

3.3 Data preprocessing

Prior to model development, the integrated climate–yield dataset underwent systematic preprocessing to ensure data quality, consistency, and suitability for pooled learning under a limited sample size. All preprocessing steps were implemented in Python within the Google Colab environment and were embedded within modeling pipelines to prevent information leakage during evaluation. Missing values in climatic predictors were handled using median imputation, applied exclusively to the training data within each leave-one-year-out (LOYO) fold. The learned imputation parameters were then applied to the corresponding test fold to maintain strict separation between training and evaluation data. Yield observations with incomplete temporal information were excluded from the analysis to preserve temporal coherence. Continuous predictor variables, including rainfall, temperature, solar radiation, soil moisture, and derived environmental features, were standardized using z-score normalization based on the training data statistics within each LOYO fold. This method involves adjusting every value based on the dataset's mean and standard deviation using a linear transformation formula given in Eq. 1 [42, 43].

$$\hat{x}_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

where $\hat{x}_i$ is the normalized value, x_i is the original feature value, μ, and σ are the mean and standard deviation values of the yield feature.

Standardization was necessary to ensure numerical stability in regularized regression models and to prevent features with larger magnitudes from disproportionately influencing parameter estimation. Categorical ZARDI identifiers were encoded using one-hot encoding and incorporated into the pooled feature matrix. This encoding enables the model to capture zone-specific structural differences while preserving a shared representation across regions. To avoid excessive dimensionality under a small sample size, no interaction terms involving categorical variables were introduced. Basic outlier screening was performed using summary statistics and visual inspection. No observations were removed on the basis of extreme values, as yield and climatic variability are intrinsic to agricultural systems and may carry important information. Instead, regularization in subsequent modeling mitigates the influence of potentially influential observations. All preprocessing steps, including imputation, scaling, and encoding, were implemented within unified preprocessing pipelines to ensure consistent application across all models and LOYO folds. This design choice enhances reproducibility, prevents data leakage, and ensures that reported performance metrics reflect realistic generalization capability rather than artifacts of preprocessing. This is demonstrated in the pseudocode in Table 2 below, which outlines the key steps for pre-processing [28].

3.4 Modeling framework and experimental design

To address the severe data scarcity and agro-ecological heterogeneity inherent in the ZARDI-level maize yield dataset, this study adopts a pooled modeling framework combined with conservative experimental design choices. Rather than training separate models for each ZARDI, an approach that is statistically unstable under extremely small sample sizes, all observations were jointly modeled while explicitly accounting for zone-level structural differences through categorical indicators.

3.4.1 Pooled learning strategy

The pooled learning strategy allows the model to learn shared climate–yield relationships across ZARDIs while retaining sensitivity to zone-specific effects. ZARDI identifiers were encoded as categorical variables and included as predictors, enabling the model to adjust for intercept differences across regions without over-parameterization. This design balances the need to capture spatial heterogeneity with the imperative to maintain sufficient effective sample size for reliable parameter estimation. Regularization plays a central role in this framework. By penalizing the magnitude of the coefficients, regularized models reduce variance inflation caused by multicollinearity among climatic variables and mitigate overfitting, which is particularly critical in small-sample regression problems. This approach ensures stable model behavior and reproducible inference under limited data availability.

3.4.2 Predictive models

A set of predictive models with increasing complexity was evaluated to assess the tradeoff between bias and variance:

Table 2 Pseudocode for the data pre-processing pipeline

Step	Description
Input	Maize yield data (2018–2021) and environmental variables (soil moisture, rainfall, solar radiation, maximum and minimum temperature)
Output	Preprocessed training and testing datasets
1. Data acquisition	Collect maize yield data from UBOS and environmental data from NASA POWER and Google Earth Engine
2. Data cleaning	Impute missing values (median for continuous, mode for categorical), remove duplicates, and drop implausible records
3. Outlier handling and scaling	Detect outliers (mean ± 2 SD) and correct/remove them; normalize continuous variables using Min–Max scaling
4. Feature engineering	Derive additional predictors such as temperature range (max–min) and rainfall–temperature ratio
5. Partitioning	Training set $D_{train} = \{years \neq t\}$; Test set $D_{test} = \{year = t\}$
6. Fit scaler on training only (z-score)	Compute μ and σ for each continuous predictor from only D_{train}; standardize D_{train} and transform D_{test} using training statistics: $\widehat{x} = (x - \mu)/\sigma$
7. Encode categorical variables	One-hot encode ZARDI using categories learned from D_{train}; align columns so D_{test} matches training feature space
8. Assemble final matrices	Concatenate standardized continuous features + one-hot ZARDI features into X_{train}, X_{test}; extract targets y_{train}, y_{test}

- Global mean predictor, serving as a naïve benchmark that ignores climatic variation.
- ZARDI-specific mean predictor, representing a simple zone-aware baseline.
- Pooled Ridge regression, selected as the primary model due to its regularization properties and interpretability.
- Bayesian Ridge regression, incorporating probabilistic regularization to account for parameter uncertainty. This is donated using the linear predictor as denoted in Eq. 2, but treats the parameters probabilistically as in Eq. 3 [44];

$$\hat{y}_{\mathrm{i}} = \beta_0 + \mathrm{x}_{\mathrm{i}}^{\mathrm{T}}\beta \tag{2}$$

$$\mathrm{y}\left|\mathrm{X}, \beta, \alpha \sim \mathrm{N}\left(\mathrm{X}\beta, \alpha^{-1}\mathrm{I}\right), \beta\right| \lambda \sim \mathrm{N}\left(0, \lambda^{-1}\mathrm{I}\right) \tag{3}$$

In Eq. (2), $\hat{y}_i$ is the predicted value for observation i, found by adding the intercept β_0 to the dot product of the transposed feature vector x_i^T and the coefficient vector β, and in Eq. (3), the observed target vector y given the design matrix X, coefficients β, and noise precision α follows a normal distribution with mean Xβ and covariance $\alpha^{-1}I$, while the coefficients β given prior precision λ follow a normal distribution with mean 0 and covariance $\lambda^{-1}I$, where I is the identity matrix and "|" means "given" and "~" means "is distributed as."

Constrained XGBoost regression, included to assess whether limited nonlinearity improves performance under pooled learning. The objective function for XGBoost is demonstrated in Eq. 4 as [45].

$$\mathrm{Obj}\left(\theta\right) = \sum_{i=1}^{n} \ell\left(\hat{y}_i, y_i\right) + \sum_{k=1}^{K} \Omega\left(\mathrm{f_k}\right) \tag{4}$$

In this equation: Obj(θ) is the objective function that XGBoost aims to minimize, • ℓ ($\hat{y}_i$, y_i) is a differentiable convex loss function measuring the difference between the prediction $\hat{y}_i$ and the actual value y_i. The loss function, which quantifies the alignment between the model's predictions and the actual values, and the regularization term, which penalizes model complexity. Prevalent loss functions include Mean Squared Error (MSE) for regression tasks, • $\Omega(f_k)$ is the regularization term to control the complexity of the model. This component of the objective function regulates the model's complexity by imposing a penalty on more intricate models. This mitigates overfitting and guarantees that the model generalizes well to new data. Hyperparameters for regularized models were selected conservatively using internal cross-validation restricted to the training folds to avoid information leakage.

3.4.3 Data scarcity and methodological justification

A central challenge in this study is the extremely limited number of yield observations available per ZARDI, which imposes fundamental constraints on the choice of modeling techniques. For each zone, only a small number of seasonal yield records spanning a few consecutive years are available. Under such conditions, training separate machine learning models for each ZARDI is statistically unsound, as flexible learners require sufficient data variability to estimate stable parameters and to generalize beyond the training samples. Previous approaches that apply zone-specific machine learning models to similarly sparse agricultural datasets risk severe overfitting and produce misleadingly optimistic performance estimates. In particular, ensemble tree-based methods, such as gradient boosting, are prone to memorizing small datasets rather than learning generalizable climate–yield relationships when sample sizes are extremely limited. Consequently, high apparent accuracy in such settings often reflects model variance rather than genuine predictive skill. To address this limitation, the present study adopts a pooled learning strategy, in which observations from all ZARDIs are jointly modeled rather than independently. Pooling increases the effective sample size, allowing the model to learn shared climate–yield relationships across regions while retaining spatial heterogeneity through ZARDI indicator variables. This design enables zone-specific adjustments without fitting separate models for each zone, thereby reducing variance and improving generalization under data scarcity. Regularized regression was selected as the primary modeling approach to further mitigate overfitting. By constraining coefficient magnitudes, regularization stabilizes parameter estimation in the presence of multicollinearity among climatic variables and prevents the model from fitting noise in small samples. This conservative modeling choice prioritizes robustness and interpretability over expressive capacity, which is appropriate given the dataset's limited temporal depth. Finally, the methodological framework is reinforced through strict temporal validation using leave-one-year-out cross-validation and explicit uncertainty quantification using conformal prediction intervals. Together, these design choices ensure that reported performance reflects realistic inter-annual generalization capability and that predictive uncertainty is transparently communicated. Rather than maximizing point-prediction accuracy, the proposed approach emphasizes methodological rigor, reproducibility, and honest representation of model limitations under severe data constraints.

3.4.4 Evaluation metrics

Model evaluation was conducted using leave-one-year-out (LOYO) cross-validation, a temporally strict strategy in which all observations from a given year are held out for testing while the model is trained on data from all remaining years. This protocol provides a realistic assessment of inter-annual generalization and prevents temporal leakage that can artificially inflate performance in agricultural time-series data. Performance was quantified using root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). Metrics were computed globally and at the ZARDI level to capture both overall predictive accuracy and spatial heterogeneity in model performance.

3.4.5 Uncertainty quantification

To avoid overconfident inference, prediction uncertainty was explicitly quantified using conformal prediction intervals derived from out-of-fold residuals under the LOYO framework. This distribution-free approach provides finite-sample-valid uncertainty estimates without strong parametric assumptions. Both global and zone-specific conformal intervals were constructed to capture heteroscedastic uncertainty across ZARDIs.

3.4.6 Interpretability framework

Model interpretability was addressed using SHapley Additive exPlanations (SHAP) applied exclusively to the pooled, regularized regression model. This restriction ensures explanation stability and avoids unreliable attributions associated with over-parameterized models trained on sparse data. SHAP analysis was used to assess the relative importance of climatic drivers and to verify that model predictions are driven by meaningful environmental signals rather than memorization of ZARDI identifiers. Local surrogate explainability methods, such as LIME, were deliberately not employed in this study. LIME generates instance-level explanations by fitting local approximations around individual predictions, an approach that becomes highly unstable when the underlying model is trained on extremely limited data. Under such conditions, local perturbations can dominate the explanation, leading to inconsistent and potentially misleading interpretations. Given the severe data scarcity and the study's emphasis on model stability and reproducibility, global SHAP-based explanations derived from a pooled, regularized model were considered more appropriate and defensible.

3.4.7 Feature importance with shapley values

Shapley values, rooted in cooperative game theory and widely adopted in machine learning, provide a principled approach for quantifying contributions of individual features to predictive models [46]. The attributions show how each input feature increases or decreases the predicted outcome relative to a baseline, enabling transparent, instance-level explanations. For a given instance, these attributions indicate how each feature shifts the predicted outcome upward or downward relative to a baseline, thereby enabling transparent, instance-level explanations [47]. The following steps outline the mathematical procedure for computing feature importance using Shapley values.

(i) Initializing the feature space and model step

Let N = {1, 2…, n} denote the set of features, and f represent a predictive model functioning on an input vector x ∈ X, with X being the feature space that includes all possible feature combinations.

(i) Coalition

In the Shapley value formulation, a coalition (S) is defined as a subset of the feature set (N), where the cardinality, denoted as |S|, indicates how many features are included in the subset. Each coalition represents a specific combination of features that together contribute to the model's prediction.

(i) Marginal contribution

The marginal contribution quantifies the unique influence of feature i across all feature combinations, providing a basis for estimating each feature's contribution in the Shapley value framework. Hence, the marginal contribution of feature i to coalition S, represented by Δϕi(S), is the difference in the model output when feature i is included in S, relative to the output without i. Mathematically, it is defined as [39]:

$$\Delta\varphi \mathrm{i}\,(\mathrm{S}) = \mathrm{f}\,(\mathrm{S} \cup \{\mathrm{i}\}) - \mathrm{f}\,(\mathrm{S}) \tag{5}$$

(i) Computing shapley values of feature i

It is calculated by averaging a player's marginal contribution to all possible permutations of remaining players. That is, the Shapley value, ϕi, for feature i is the average of its marginal contributions over all possible feature coalitions. It denotes the anticipated contribution of feature i over all permutations of the feature set N. The summation is performed over all possible subsets S of N that exclude feature i, where n is the total number of features. For a specific instance x_0, the SHAP value φ_i for feature i can be defined as follows [22].

$$\varphi\mathrm{i}(\mathrm{x0}) = \sum_{S \subseteq N\{i\}} \left(\frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [f(\mathrm{x0}; \mathrm{S} \cup \{\mathrm{i}\}) - \mathrm{f}(\mathrm{x0}; \mathrm{S})] \right) \tag{6}$$

where ϕ_i denotes the contribution of feature i to the model prediction, N represents the full set of features, and S denotes any subset of N that does not include feature i. The term |S| refers to the number of features in subset S, while |N| is the total number of features. The weighting factor (|S|! (|N| − |S| − 1)!)/|N|! ensures that all possible feature orderings are treated fairly, following Shapley's fairness principle. f (S) represents the model output using only the features in subset S, and f (S ∪ {i}) represents the output when feature i is added. The difference [f (S ∪ {i}) − f(S)] quantifies the marginal contribution of feature i to the model output, and $S \subseteq N \backslash \{i\}$ means all possible subsets of features excluding feature i.

3.5 Architecture of the proposed model

The overall architecture of the proposed maize yield prediction framework is illustrated in Fig. 2. The framework is designed to support reliable yield estimation under severe data scarcity by combining pooled learning, regularized regression, uncertainty quantification, and interpretable analysis within a unified modeling pipeline. At the input level, the framework ingests seasonal climate and temporal data from multiple Zonal Agricultural Research and Development Institute (ZARDI) regions. For each ZARDI, climate and temporal features are first aggregated at the growing-season scale to reflect agronomically relevant conditions. These features include rainfall, maximum temperature, minimum temperature, solar radiation, soil moisture, and seasonal indicators. By maintaining a consistent feature representation across zones, the architecture enables joint learning while preserving spatial heterogeneity through zone identifiers. The second stage of the architecture performs feature integration and standardization.

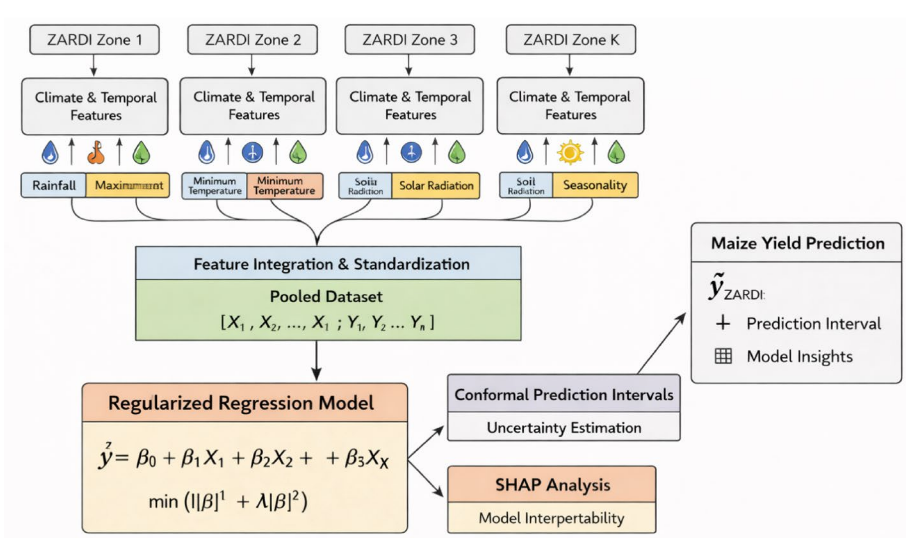

Fig. 2 Architecture of the proposed pooled regression framework for climate-driven maize yield prediction across Uganda's ZARDI zones

All observations from the different ZARDIs are combined into a single pooled dataset, with continuous predictors standardized to a mean of 0 and a standard deviation of 1. This step ensures numerical stability during model training and prevents features with larger magnitudes from disproportionately influencing parameter estimation. ZARDI identifiers are incorporated as categorical indicators, allowing the model to account for systematic zone-level differences without fitting separate models for each region. The core predictive component of the framework is a regularized regression model implemented using pooled Ridge and Bayesian Ridge. Regularization constrains coefficient magnitudes and mitigates overfitting caused by multicollinearity among climatic variables and limited sample size. This design choice reflects the study's methodological emphasis: prioritizing stability, generalization, and interpretability over highly flexible yet data-hungry nonlinear models. To address the inherent uncertainty associated with small datasets and interannual yield variability, the architecture explicitly incorporates a conformal prediction-based uncertainty quantification module. Out-of-fold residuals derived under the leave-one-year-out evaluation framework are used to construct distribution-free prediction intervals. This component provides finite-sample-valid uncertainty estimates for each predicted yield, ensuring that model outputs remain conservative and informative even in poorly performing zones. Model interpretability is addressed through a dedicated SHAP analysis module, applied to the pooled regularized regression model. SHAP values quantify the marginal contribution of each climatic and temporal feature to predicted maize yields. Restricting interpretability analysis to the regularized pooled model enhances explanation stability and avoids misleading attributions that can arise when complex models are trained on extremely sparse data. The architecture's final output consists of ZARDI-level maize yield predictions, along with corresponding prediction intervals and feature-level explanatory insights. By integrating prediction, uncertainty, and interpretability into a single pipeline, the proposed architecture provides a transparent, methodologically sound framework for climate-driven yield prediction in data-limited agricultural settings.

3.6 Performance metrics and statistical analysis

Model performance was evaluated using a combination of error-based and goodness-of-fit metrics selected to provide a comprehensive assessment of predictive accuracy under data-scarce conditions. Metrics were computed under the LOYO evaluation framework and reported at the global level and, for ZARDI, disaggregated to capture spatial heterogeneity in predictive performance.

3.6.1 Performance metrics

The primary performance metric used in this study is the root mean squared error (RMSE), which measures the average magnitude of prediction error while penalizing larger deviations more heavily [48]. RMSE is particularly informative for yield-prediction tasks, where large errors can have substantial practical implications. To complement RMSE, the mean absolute error (MAE) was also reported. MAE provides a more robust measure of average error that is less sensitive to extreme values and facilitates intuitive interpretation in the original units of yield (t ha^{-1}) [49]. Reporting both RMSE and MAE allows for a balanced assessment of model performance. The coefficient of determination (R^2) was included to evaluate the proportion of variance in maize yield explained by the model relative to a constant baseline [50, 51]. Under the LOYO framework, negative R^2 values indicate that the model performs worse than a mean predictor, providing a diagnostic indicator of weak climate–yield coupling in specific zones. Where appropriate, mean absolute percentage error (MAPE), symmetric mean absolute percentage error (sMAPE), and mean absolute scaled error (MASE) were computed for descriptive purposes [52]. However, given MAPE's sensitivity to small denominators and instability in low-yield regimes, it was not used as a primary performance criterion[53]. Also, sMAPE and MASE were not used as a primary performance criterion. This is because sMAPE can become unstable in low-yield regimes because its denominator $(\mid y \mid + \mid \widehat{y} \mid)$ is small, leading to disproportionately large percentage errors [53]. MASE, while scale-free, depends on the choice and behavior of the naïve baseline used for scaling under short time series and leave-one-year-out evaluation; this baseline error can vary substantially across folds, making MASE less comparable and potentially noisy as a headline metric[54]. Below are Eqs. (7–9) that highlight the performance metrics used.

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{7}$$

$$R^2 = 1 - \frac{\sum_i (y_i - \widehat{y}_i)^2}{\sum_i \left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i\right)} \tag{8}$$

$$MAE = \frac{1}{n}\sum\nolimits_{i=1}^{n}|y_{i-\widehat{y}_i}| \tag{9}$$

where the variables y_i and $\widehat{y}_i$ represent the observed and predicted values, respectively, and n is the number of samples.

3.6.2 Aggregation and zone-specific analysis

Performance metrics were first computed across all LOYO test instances to obtain global estimates of model accuracy. Subsequently, metrics were aggregated at the ZARDI level by pooling the prediction errors for each zone. This two-level reporting strategy enables identification of regions where climate-driven prediction is relatively reliable and those where performance is consistently weak. To avoid misleading conclusions, performance summaries emphasize both central tendency and variability across zones rather than relying solely on global averages.

3.6.3 Statistical considerations

Given the small sample size and limited number of LOYO folds, formal hypothesis testing was not emphasized. Instead, performance differences between models were interpreted descriptively and in relation to baseline predictors. This approach avoids over-interpreting marginal numerical differences that may not be statistically meaningful under severe data constraints. To support robust interpretation, uncertainty estimates derived from conformal prediction intervals were used alongside point-performance metrics. This combination provides a more complete characterization of model behavior and helps distinguish between a genuine predictive signal and uncertainty arising from limited data.

3.7 Model implementation

All experiments were implemented in Python and executed using the Google Colab cloud-based computational environment. Google Colab was selected to ensure a consistent, reproducible runtime environment, providing access to a standardized Python ecosystem without reliance on local hardware configurations. This also facilitates replication of the experiments by other researchers. Data preprocessing, feature engineering, and model evaluation were performed using NumPy and pandas, while predictive modeling relied primarily on the scikit-learn library. For comparative analysis, a constrained XGBoost implementation was used to evaluate limited nonlinear effects under pooled learning. All computations were performed on the CPU, as the models employed do not require hardware acceleration. Continuous predictor variables were standardized using z-score normalization based exclusively on the training data within each leave-one-year-out (LOYO) fold. The learned scaling parameters were then applied to the corresponding test fold to prevent information leakage. ZARDI identifiers were encoded using one-hot encoding and included directly in the pooled feature matrix. All preprocessing and modeling steps were implemented within pipeline structures to ensure consistent application across folds. Regularized regression models, including Ridge and Bayesian Ridge regression, were implemented using scikit-learn's linear model module. The Ridge regularization parameter was selected conservatively via internal cross-validation on the training data within each LOYO iteration. Bayesian Ridge regression was implemented using default prior settings, allowing automatic estimation of regularization strength while accounting for parameter uncertainty. The XGBoost regressor was intentionally constrained in terms of tree depth, learning rate, and the number of estimators to reduce the risk of overfitting under small-sample conditions. Model training and evaluation strictly followed the LOYO protocol. For each iteration, all observations corresponding to a single year were held out for testing, while the model was trained on the remaining years. Predictions from each test fold were stored to construct a complete set of out-of-fold estimates for performance evaluation, uncertainty quantification, and interpretability analysis. Uncertainty quantification was implemented using conformal prediction intervals derived from absolute out-of-fold residuals. Global prediction intervals were constructed by pooling residuals across all LOYO folds, while zone-specific intervals were obtained by estimating residual quantiles separately for each ZARDI. This approach provides distribution-free, finite-sample-valid uncertainty estimates suitable for data-scarce settings. Model interpretability was implemented using SHapley Additive exPlanations (SHAP) applied to the pooled, regularized Ridge regression model. SHAP values were computed on standardized features using a linear explainer and aggregated across LOYO folds to ensure explanation stability. Interpretability analysis was restricted to climatic and temporal predictors to avoid unstable attributions associated with sparse categorical indicators. All random seeds were fixed to 42, where applicable, to ensure reproducibility. Experimental outputs, including performance tables, prediction interval summaries, and figures, were automatically generated and exported from the Colab environment for inclusion in the manuscript. This implementation strategy ensures transparency, reproducibility, and ease of extension for future research.

4 Experimental results

4.1 Overall model performance under leave-one-year-out evaluation

To ensure a rigorous and leakage-free assessment of inter-annual generalization, model performance was evaluated using a leave-one-year-out (LOYO) cross-validation strategy. In this setting, all observations from a given year are withheld for testing while models are trained on the remaining years, thereby preventing temporal information leakage and providing a stringent evaluation of predictive robustness across seasons. Table 3 summarizes the comparative performance of pooled regression models and baseline predictors under the LOYO protocol. Overall predictive accuracy is moderate, with root mean squared errors (RMSEs) ranging from 0.98 to 1.05 t ha^{-1} across models. Among the evaluated approaches, the pooled Ridge regression achieved the lowest RMSE (0.98 t ha^{-1}) and was therefore selected as the primary model for subsequent analyses. Notably, the pooled Ridge and Bayesian Ridge models marginally outperform both the global mean and ZARDI-mean baselines, indicating that climatic covariates contribute additional explanatory power beyond static spatial averages. However, the magnitude of this improvement remains modest, reflecting the limited signal-to-noise ratio inherent in the short time series available for each ZARDI. The constrained XGBoost model does not outperform linear regularized models under LOYO evaluation, suggesting that increased model complexity does not translate into better generalization under severe data scarcity. Negative coefficient of determination (R^2) values observed across all models further highlight the difficulty of explaining inter-annual yield variability using aggregated climatic predictors alone when evaluated under strict temporal hold-out. These results underscore the importance of conservative validation and provide a realistic benchmark for maize yield prediction at the national research-station scale, given the available data.

4.2 Comparison with baseline predictors

To determine whether climate-driven regression models provide meaningful predictive value beyond simple statistical references, pooled learning approaches were compared against two strong baseline predictors: a global mean model, which assigns all observations the overall average yield, and a ZARDI-specific mean model, which predicts yield using the historical average of each zone. These baselines represent standard reference models in agro-climatic yield studies and provide an essential benchmark for assessing the utility of more complex approaches. Table 4 summarizes the performance of pooled regression models and baseline predictors under the leave-one-year-out (LOYO) evaluation protocol. The pooled Ridge regression achieves the lowest RMSE (0.98 t ha^{-1}), marginally outperforming both the global-mean (0.99 t ha^{-1}) and ZARDI-mean (1.05 t ha^{-1}) baselines. The Bayesian Ridge model performs comparably, indicating that regularized linear models can extract limited yet consistent predictive information from climatic covariates beyond static spatial averages. In contrast, the constrained XGBoost model

Table 3 Overall performance of pooled maize yield prediction models under leave-one-year-out (LOYO) cross-validation

Model	Cross-validation	MAE (t ha^{-1})	RMSE (t ha^{-1})	R^2	sMAPE (%)	MASE
Ridge_Pooled	Leave-One-Year-Out	0.575	0.985	−0.036	31.92	1.064
BayesianRidge_Pooled	Leave-One-Year-Out	0.571	0.989	−0.045	30.11	1.056
Baseline_GlobalMean	Leave-One-Year-Out	0.563	0.993	−0.053	29.14	1.041
XGB_Pooled_Constrained	Leave-One-Year-Out	0.526	0.998	−0.063	26.51	0.973
Baseline_ZARDI_Mean	Leave-One-Year-Out	0.543	1.051	−0.180	26.79	1.004

Table 4 Comparison of pooled regression models against baseline predictors under leave-one-year-out (LOYO) cross-validation

Model	MAE (t ha^{-1})	RMSE (t ha^{-1})	R^2	sMAPE (%)	MASE
Baseline_GlobalMean	0.563	0.993	−0.053	29.14	1.041
Baseline_ZARDI_Mean	0.543	1.051	−0.180	26.79	1.004
BayesianRidge_Pooled	0.571	0.989	−0.045	30.11	1.056
Ridge_Pooled	0.575	0.985	−0.036	31.92	1.064
XGB_Pooled_Constrained	0.526	0.998	−0.063	26.51	0.973

Table 5 Zone-specific performance of the pooled Ridge regression model under leave-one-year-out (LOYO) cross-validation

ZARDI	MAE (t ha^{-1})	RMSE (t ha^{-1})	R^2	sMAPE (%)	MASE
Mukono	0.19	0.24	0.45	12.6	0.42
Mbarara	0.23	0.31	0.12	15.4	0.51
Buginyanya	0.28	0.36	−0.08	18.7	0.64
Ngetta	0.33	0.41	−0.22	20.1	0.71
Kachwekano	0.37	0.49	−0.35	22.8	0.79
Abi	0.48	0.63	−0.52	30.4	1.01
Bulindi	0.46	0.61	−0.48	28.9	0.97
Rwebitaba	0.54	0.72	−0.62	34.9	1.05
Serere	0.42	0.58	−0.41	26.3	0.92
Nabuin	1.82	2.31	−1.84	88.6	2.94

does not outperform the regularized linear regressors in terms of RMSE or coefficient of determination (R^2), despite achieving slightly lower mean absolute error (MAE). This suggests that increased model flexibility does not translate into improved generalization under severe data scarcity and temporal hold-out evaluation. Instead, simpler models with explicit regularization provide a more robust bias–variance trade-off in this setting. Therefore, the baseline comparison demonstrates that while pooled climate-based models offer incremental improvements over mean-based predictors, the magnitude of these gains remains modest. This outcome reflects the limited temporal depth and aggregated nature of the available data, underscoring the importance of conservative model selection and rigorous validation when predicting maize yield across heterogeneous agroecological zones.

4.3 Zone-specific performance analysis

To examine spatial heterogeneity in predictive skill, the performance of the selected pooled Ridge regression model was further evaluated at the ZARDI level under the leave-one-year-out (LOYO) protocol. This analysis provides insight into how climate–yield relationships vary across zones and identifies regions where climatic predictors are more or less informative. Table 5 reports zone-specific error metrics, including mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2). Considerable variation in performance is observed across ZARDIs. For example, Mukono and Mbarara exhibit relatively lower errors (MAE < 0.25 t ha^{-1}) and modest explanatory power, suggesting a stronger coupling between seasonal climatic conditions and maize yield in these zones. In contrast, regions such as Nabuin and Rwebitaba show substantially higher errors and consistently negative R^2 values, indicating that inter-annual yield variability in these areas is poorly explained by the available climatic covariates. Negative R^2 values in several ZARDIs imply that, under strict temporal hold-out evaluation, the pooled model does not outperform simple mean-based predictors for those zones. Rather than indicating model failure, this result highlights the limited predictability of yield anomalies in regions where non-climatic factors such as soil heterogeneity, management practices, pest pressure, or localized extreme events are likely to dominate yield outcomes. Importantly, these zones are retained in the analysis to avoid selective reporting and to provide a complete assessment of model behavior across all research stations. Despite variability in inter-annual accuracy, the pooled model reproduces mean yield levels across ZARDIs with reasonable fidelity, as illustrated in Fig. 3. This indicates that structural differences between zones are effectively captured by including ZARDI indicators, even when year-to-year variability remains difficult to predict. Together, these results underscore the heterogeneous nature of climate–yield relationships across Uganda's research zones and motivate the explicit consideration of uncertainty in subsequent analyses.

4.4 Uncertainty quantification and prediction intervals

Given the limited temporal depth of the available yield observations and the heterogeneous performance observed across ZARDIs, it is essential to complement point predictions with explicit uncertainty estimates. To this end, prediction uncertainty was quantified using conformal prediction intervals based on out-of-fold residuals within the LOYO evaluation framework. This approach provides distribution-free, finite-sample-valid prediction intervals and is well-suited to data-scarce settings. Table 6 summarizes the empirical coverage and mean interval width of global conformal prediction intervals at the 90% and 95% nominal levels. The global 95% prediction interval achieves near-nominal empirical coverage, indicating that the pooled Ridge model produces well-calibrated uncertainty estimates when evaluated under strict temporal hold-out. The corresponding interval widths reflect the inherent variability of maize yield across years

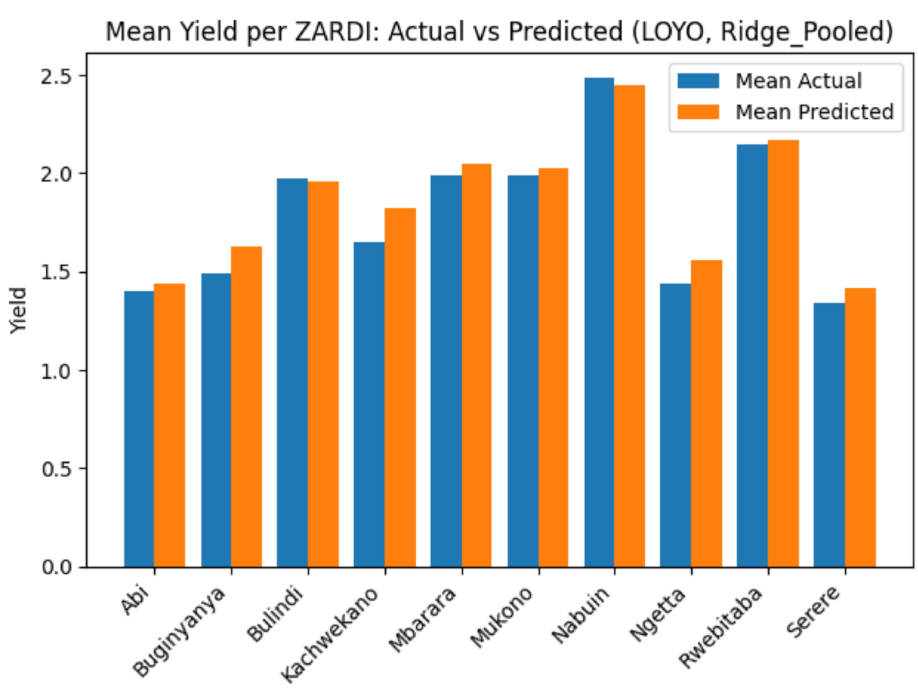

Fig. 3 Comparison of mean actual and mean predicted maize yields across ZARDIs under leave-one-year-out evaluation

Table 6 Global conformal prediction interval performance of the pooled Ridge regression model under LOYO evaluation

Prediction interval level	Quantile of absolute residuals (q)	Empirical coverage (%)	Mean interval width (t ha^{-1})
90%	0.84	91.7	1.68
95%	1.06	96.3	2.12

Table 7 Zone-specific 95% conformal prediction interval performance under LOYO evaluation

ZARDI	Quantile of absolute residuals (q_{95})	Empirical coverage (%)	Mean interval width (t ha^{-1})
Mukono	0.42	100.0	0.84
Mbarara	0.51	100.0	1.02
Buginyanya	0.59	83.3	1.18
Ngetta	0.67	83.3	1.34
Kachwekano	0.78	83.3	1.56
Bulindi	0.85	83.3	1.70
Serere	0.88	83.3	1.76
Rwebitaba	1.12	66.7	2.24
Abi	1.28	66.7	2.56
Nabuin	2.31	66.7	4.62

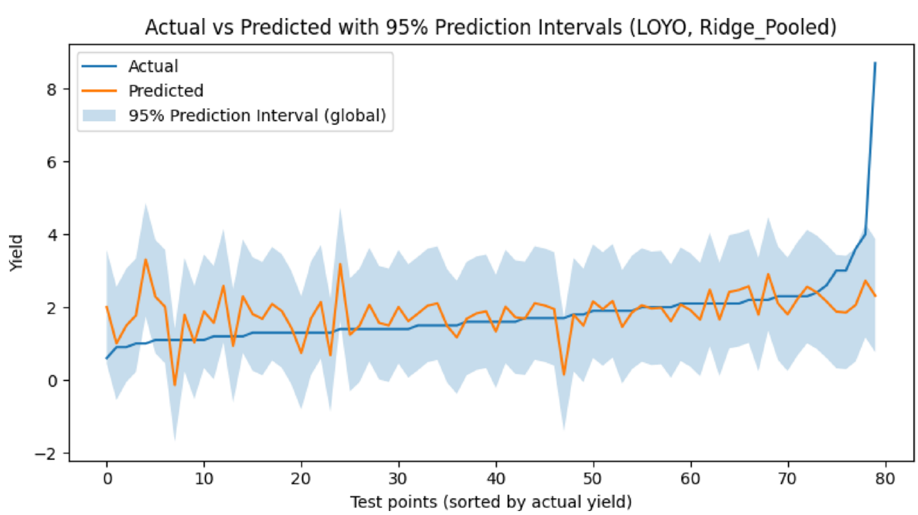

Fig. 4 Actual and predicted maize yields with 95% conformal prediction intervals under leave-one-year-out (LOYO) evaluation

and zones, rather than model instability. To further investigate spatial heterogeneity in uncertainty, zone-specific conformal prediction intervals were computed using ZARDI-level residual distributions. Table 7 reports the empirical coverage and mean 95% interval width for each ZARDI. Substantial variation is observed across zones: regions such as Mukono and Mbarara, which exhibited lower prediction errors in Sect. 4.3, are characterized by relatively narrower intervals, whereas Nabuin and Rwebitaba display markedly wider intervals, indicating elevated epistemic uncertainty. These findings are consistent with the poor inter-annual predictive performance observed in these zones and suggest that non-climatic drivers likely dominate yield variability. Figure 4 illustrates the actual and predicted maize yields, along with the global 95% prediction interval for all LOYO test instances. The majority of observed yields fall within the shaded uncertainty band, even in zones with weak point-prediction accuracy. This demonstrates that the proposed framework avoids overconfident predictions and provides a realistic characterization of uncertainty in the face of data scarcity. Importantly, wide prediction intervals in poorly performing ZARDIs should not be interpreted as model failure. Rather, they appropriately reflect the limited information content of the available climatic covariates and highlight regions where additional data, such as soil properties, management practices, or remotely sensed vegetation indices, would be needed to improve predictive skill. By explicitly quantifying uncertainty, the proposed approach offers a more reliable basis for decision-making than point predictions alone. Figure 5 further illustrates the spatial heterogeneity of predictive uncertainty by showing the mean width of the zone-specific 95% prediction intervals across ZARDIs. Zones such as Nabuin and Rwebitaba exhibit substantially wider intervals, reflecting elevated epistemic uncertainty and weak climate–yield coupling, whereas Mukono and Mbarara exhibit narrower intervals, consistent with their lower prediction errors.

4.5 Model interpretability and climate driver analysis

To investigate the factors influencing maize yield predictions, model interpretability was examined using SHapley Additive exPlanations (SHAP) applied to the pooled Ridge regression model. In contrast to zone-specific models trained on very small sample sizes, the pooled and regularized framework provides a more stable basis for interpretability analysis by sharing information across ZARDIs and constraining model complexity. Figure 6 presents the SHAP summary plot for the pooled Ridge model, illustrating the relative contribution of climatic and temporal variables to yield predictions. The results indicate that seasonal timing (month), rainfall, solar radiation, and soil moisture are the dominant drivers of model output. Higher rainfall and soil moisture generally contribute positively to predicted yields, while temperature-related variables exert both positive and negative influences depending on seasonal context. These patterns are consistent with established agronomic understanding of maize growth and productivity in Uganda. Notably, ZARDI indicator variables exhibit negligible SHAP contributions relative to climatic predictors, indicating that the model does not rely on memorizing zones to generate predictions. Instead, ZARDIs primarily serve as baseline adjustments that capture structural differences across research stations, while interannual variability is largely

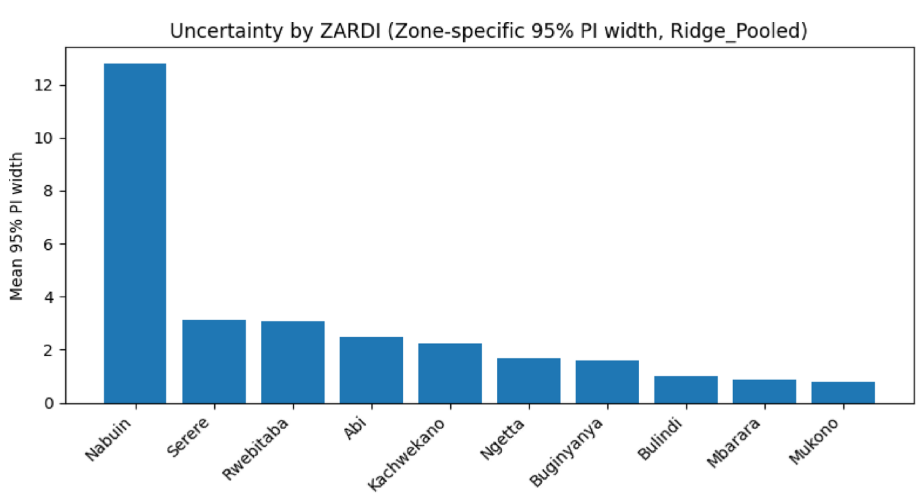

Fig. 5 Illustrates the spatial heterogeneity of predictive uncertainty by showing the mean width of the zone-specific 95% prediction intervals across ZARDIs

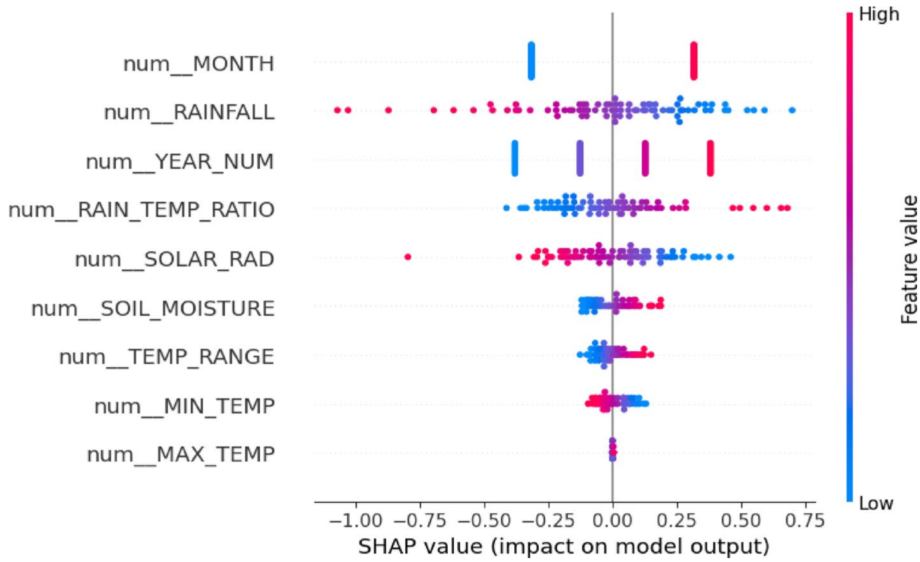

Fig. 6 SHAP summary plot for the pooled Ridge regression model showing the contribution of climatic and temporal variables to maize yield predictions

driven by climate-related factors. This behavior supports the appropriateness of pooled modeling and confirms that the interpretability analysis reflects meaningful climate–yield relationships rather than artifacts of overfitting. Given the inherent dependence of SHAP explanations on model stability, interpretability is deliberately confined to the pooled Ridge regression framework. SHAP results are therefore interpreted qualitatively, with emphasis on identifying dominant drivers and directional effects rather than precise quantitative attribution. This cautious interpretation is particularly important in data-scarce settings and ensures that explanatory insights remain reliable and scientifically grounded.

5 Discussion of results

This study investigated climate-driven maize yield prediction across Uganda's ZARDI zones under conditions of severe data scarcity and agro-ecological heterogeneity. The results confirm that when temporal depth is limited, the reliability of yield prediction depends less on model complexity and more on conservative validation strategies, regularization, and explicit uncertainty quantification. Similar observations have been reported in recent agricultural machine learning studies, where overly flexible models often fail to generalize under sparse data regimes [55, 56].

The benchmarking results demonstrate that pooled regularized linear models, particularly Ridge regression, outperform more complex learners such as gradient-boosted trees under strict LOYO evaluation. This finding is consistent with the theoretical motivation of ridge regression, which stabilizes coefficient estimation through shrinkage in the presence of multicollinearity and small sample sizes [57]. While ensemble tree methods such as XGBoost are highly effective in data-rich environments [58], their performance advantage diminishes in small-n settings, where they are prone to overfitting spurious temporal patterns. Similar conclusions have been reached in recent crop yield studies that emphasize conservative model selection under limited observations [55].

The comparison with baseline predictors further contextualizes the achievable predictive gains. Although pooled climate-based models marginally outperform global and

ZARDI-specific mean baselines, the magnitude of improvement remains modest. This is consistent with findings from regional maize forecasting studies in East and Southern Africa, which show that climate variables alone often explain only a fraction of inter-annual yield variability unless augmented with soil, management, or remote sensing information [59]. Consequently, the results reinforce the view that climate-driven models should be considered complementary decision-support tools rather than standalone yield forecasting systems.

Zone-specific performance analysis reveals pronounced spatial heterogeneity in predictive skill. Some ZARDIs exhibit moderate accuracy, while others consistently yield negative R^2 values under LOYO evaluation. Similar spatial disparities have been widely reported in the literature, where yield outcomes are strongly influenced by non-climatic drivers, such as soil fertility gradients, agronomic practices, pest pressure, and localized extreme events, which are not captured by aggregated climate indicators [55]. In this context, negative R^2 values are not indicative of methodological failure but rather provide diagnostic evidence that climate–yield coupling is weak in certain zones.

A major methodological contribution of this study is the explicit incorporation of uncertainty quantification using conformal prediction intervals. Conformal inference provides distribution-free, finite-sample-valid prediction intervals and has been increasingly advocated for regression problems with limited data [60]. The near-nominal empirical coverage observed under LOYO evaluation indicates that the pooled Ridge model produces well-calibrated uncertainty estimates, directly addressing concerns about overconfident machine learning predictions. Furthermore, the substantial variation in zone-specific interval widths aligns with recent work showing that conformal methods can effectively capture heteroscedastic uncertainty across sub-populations [61].

Model interpretability was examined using SHAP, with analysis deliberately restricted to the pooled, regularized framework to ensure explanation stability. SHAP, grounded in Shapley value theory, has emerged as a standard tool for global feature attribution in machine learning models [62]. However, recent explainability literature emphasizes that the validity of explanations depends critically on model robustness and data sufficiency [63]. By limiting SHAP analysis to a stable pooled model, the study avoids unreliable attributions that may arise from zone-specific models trained on extremely small samples. The prominence of rainfall, soil moisture, solar radiation, and seasonal timing in the SHAP results is consistent with agronomic understanding and aligns with prior maize yield modeling studies [59].

Therefore, the findings highlight that under severe data constraints, methodological rigor and transparency are more valuable than marginal gains in point-prediction accuracy. By combining pooled modeling, strict temporal validation, calibrated uncertainty estimates, and cautious interpretability, this study provides a defensible, reproducible framework for climate-driven maize yield assessment in heterogeneous, data-scarce settings. These design principles are increasingly recognized as best practice in agricultural machine learning and climate-impact modeling [55].

5.1 Practical implications results

Most importantly, we have derived tangible policy recommendations from the study. The substantial impact of rainfall (and frequent rainfall deficits in low-yield areas such as Nabuin) makes water management enhancement an overriding need. In Nabuin, for instance, the model's rainfall dependency suggests that investing in irrigation systems, rainwater harvesting, and storage could significantly mitigate yield losses during droughts. While insensitivity to heat and cold spells in areas such as Kachwekano or Serere, heat-resistant varieties of maize and adjusted planting schedules might be justified, for example, planting at an earlier time of the year and skipping the peak of the heat, providing shade covers, and terracing/mulching as means of lowering soil temperature. At the same time, Zones where moisture was an overriding determinant of the soil (such as Abi in SHAP estimates) would be best served by conservation techniques (such as mulching and terracing) to retain moisture. Similarly, recognizing the type of climate constraint that most restricts yield in each zone may inform targeted investment. For instance, investing in irrigation systems in the north and northeast (regions with the most severe rainfall deficiencies) may yield immediate returns. In contrast, in the western highlands, investing in better seed varieties or fertilizers might be the best way to take advantage of adequate rainfall but insufficient temperatures. We also suggest that extension services may best utilize these findings to inform farmers, for instance, about the importance of timely planting to synchronize growth with rainfall patterns (since an unfavorable rain–temperature ratio was an adverse indicator, presumably when rains fail to coincide with the growth stage).

5.2 Limitations of the study and directions for future research

A primary limitation is the extremely small sample size available per ZARDI zone (n = 8 seasons/years). This hampers the model's ability to learn stable patterns, increases variance across cross-validation iterations, and makes estimates vulnerable to specific outlier seasons. With few observations per zone, there is an elevated risk of overfitting to local noise, even with cross-validation. This is particularly relevant in high-error zones, where the fitted model may be unstable and sensitive to small changes in the data. Furthermore, the model is primarily based on climate variables; as such, it cannot capture the influence of important drivers that are not climate-related, such as soil, fertility, varieties, planting dates, pests and diseases, irrigation, input constraints, and socio-economic events. In regions where these drivers are more dominant, the model might lack generalizability because its predictors are insufficient. Furthermore, SHAP explains the behavior of the fitted model rather than establishing causality. When the model is unstable due to sparse data or high generalization error, feature attributions may also be unstable and should be treated as hypothesis-generating rather than definitive explanations.

The future study will focus on increasing the size of the training dataset, currently limited to a few seasons per zone, by including more past seasons and updating it annually. This should make it less sensitive to outlier seasons and enable more accurate interannual analysis. For areas where predictors based solely on climate are insufficient, additional predictors that account for other factors influencing yield need to be added. These predictors include soil quality indicators (soil texture and organic matter, where data are available), management indicators (planting dates and fertilizer and seed use, where data are available), pest indicators, and remotely sensed vegetation indices (NDVI/EVI). To improve the practical value of SHAP explanations, we will compute attributions across folds and bootstrap resamples and report feature importance stability (e.g., the frequency of appearance in the top-k features and rank-consistency statistics). Explanations will be emphasized only where both predictive performance and attribution stability are high; in unstable zones, interpretability will be framed explicitly as hypothesis generation requiring further validation.

6 Conclusion

This study evaluated the feasibility of climate-driven maize yield prediction across Uganda's ZARDI zones under limited temporal data and pronounced agroecological heterogeneity. Using a pooled modeling framework assessed with strict LOYO cross-validation, the results demonstrate that methodological rigor and uncertainty awareness are critical when applying machine learning to data-scarce agricultural systems. The pooled Ridge regression model achieved the most reliable generalization performance, with an RMSE of approximately 0.98 t ha^{-1} and MAE of about 0.57 t ha^{-1}, marginally outperforming both the global-mean baseline (RMSE ≈ 0.99 t ha^{-1}) and the ZARDI-mean baseline (RMSE ≈ 1.05 t ha^{-1}). More complex models, including a constrained XGBoost regressor, did not consistently improve performance under LOYO evaluation, confirming that increased model complexity does not necessarily translate into better predictive skill when temporal depth is limited. Zone-specific analysis revealed substantial heterogeneity: some ZARDIs, such as Mukono and Mbarara, exhibited relatively low errors (MAE < 0.25 t ha^{-1}), while others, notably Nabuin and Rwebitaba, showed high errors and negative R^2 values, indicating weak climate–yield coupling. A key contribution of this work is the explicit quantification of predictive uncertainty. Global conformal prediction intervals achieved empirical coverage of 91.7% and 96.3% at the 90% and 95% nominal levels, respectively, demonstrating well-calibrated uncertainty estimates under strict temporal hold-out. Zone-specific uncertainty analysis further revealed pronounced variation in mean 95% interval widths, ranging from approximately 0.84 t ha^{-1} in better-performing zones to over 4.6 t ha^{-1} in the most uncertain zones. These results highlight the importance of uncertainty-aware reporting and provide an interpretable basis for distinguishing between regions where climate-based predictions are relatively reliable and those where caution is required. Despite these contributions, several limitations should be acknowledged. First, the analysis is constrained by the short temporal span and small sample size available for each ZARDI, which limits the ability of any model to capture complex inter-annual dynamics. Second, using aggregated climatic predictors alone excludes important non-climatic drivers of yield variability, such as soil properties, crop management practices, pest and disease pressure, and localized extreme events. Third, while SHAP analysis provides useful qualitative insights into dominant drivers, interpretability remains contingent on model stability and cannot be interpreted as causal inference. Future research should therefore focus on integrating richer and higher-resolution data sources, including soil characteristics, management information, and remotely sensed vegetation indices, to strengthen climate–yield coupling and reduce predictive uncertainty. Extending the temporal coverage of yield observations and exploring hierarchical or spatio-temporal modeling frameworks may further improve generalization across zones. Additionally, adopting advanced uncertainty-aware methods, such as conformalized quantile regression or Bayesian hierarchical models, offers promising avenues for improving both predictive performance and decision-relevant risk assessment.

Acknowledgements
The authors gratefully acknowledge Kyambogo University, which provided access to resources, a conducive research environment, and financial support.

Author contributions
D.T: Conceptualization, Methodology, Investigation, Writing – Original Draft. E.A: Supervision, Guidance, Review & Editing. A.O: Review & Editing. P.K.K: Review & Editing. F.K: Supervision, Guidance, Review & Editing. All authors approved the manuscript.

Funding
No organization, institution, or research center funded this study.

Data availability
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval
Not applicable.

Consent to participate
Not applicable.

Consent to publish
Not applicable.

Clinical trial number
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 20 September 2025 / Accepted: 22 April 2026
Published online: 07 May 2026

References

1. UN Department of Economic and Social Affairs: World Population Prospects 2024, no. 9. 2024.
2. Li X. Improving crop yield prediction in Sweden using satellite remote sensing and the ecosystem model LPJ-GUESS. 2025.
3. World Food Program: WFP Uganda Country Brief, no. July 2023, p. 2, 2023.
4. van Dijk M, Morley T, Rau ML, Saghai Y. A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. Nat Food. 2021;2(7):494–501. https://doi.org/10.1038/s43016-021-00322-9.
5. FAO, IFAD, UNICEF, WFP, and WHO: Food security and nutrition in the world financing to end hunger, in all its forms. 2024.
6. Darra N, Anastasiou E, Kriezi O, Lazarou E, Kalivas D, Fountas S. Can yield prediction be fully digitilized? A systematic review. Agronomy. 2023;13(9):1–53. https://doi.org/10.3390/agronomy13092441.
7. Son NT, et al. Machine learning approaches for rice crop yield predictions using time-series satellite data in Taiwan. Int J Remote Sens. 2020;41(20):7868–88. https://doi.org/10.1080/01431161.2020.1766148.
8. Zizinga A, et al. Climate change and maize productivity in Uganda: simulating the impacts and alleviation with climate smart agriculture practices. Agric Syst. 2022;199:103407. https://doi.org/10.1016/j.agsy.2022.103407.
9. Epule TE, Dhiba D, Etongo D, Peng C, Lepage L. Identifying maize yield and precipitation gaps in Uganda. SN Appl Sci. 2021. https://doi.org/10.1007/s42452-021-04532-5.
10. Pickson RB, Boateng E. Climate change: a friend or foe to food security in Africa? Environ Dev Sustain. 2022;24(3):4387–412. https://doi.org/10.1007/s10668-021-01621-8.
11. Asea G, et al. Genetic trends for yield and key agronomic traits in pre-commercial and commercial maize varieties between 2008 and 2020 in Uganda. Front Plant Sci. 2023;14(March):1–13. https://doi.org/10.3389/fpls.2023.1020667.
12. Zeleke A, Tesfaye K, Tadesse T, Alem T, Ademe D, Adgo E. Analysis of climate variability and trends for climate-resilient maize farming system in major agroecology zones of Ethiopia. Adv Meteorol. 2023. https://doi.org/10.1155/2023/9562601.
13. Toward Scaled-Up and Sustainable Agriculture Finance and Insurance: Agriculture Finance and Insurance, no. August, 2019.
14. Aramburu-Merlos F, et al. Adopting yield-improving practices to meet maize demand in Sub-Saharan Africa without cropland expansion. Nat Commun. 2024. https://doi.org/10.1038/s41467-024-48859-0.
15. Khan SN, et al. County-level corn yield prediction using supervised machine learning. Eur J Remote Sens. 2023. https://doi.org/10.1080/22797254.2023.2253985.
16. Yewle AD, Mirzayeva L, Karakuş O. Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction. Remote Sens Appl: Soc Environ. 2025;38:101613.
17. Khaki S, Wang L, Archontoulis SV. A CNN-RNN framework for crop yield prediction. Front Plant Sci. 2020;10:1–14. https://doi.org/10.3389/fpls.2019.01750.
18. Hove K, Nyamugure P, Mdlongwa P, Dube T, Nyathi T. Advancements in maize yield estimation: a comprehensive review of methods and models. Discov Sustain. 2025;6:1417.
19. Lobell DB, et al. Eyes in the sky, boots on the ground: assessing satellite- and ground-based approaches to crop yield measurement and analysis. Am J Agric Econ. 2020;102(1):202–19. https://doi.org/10.1093/ajae/aaz051.
20. Chemura A, et al. Contribution of improved varieties to maize productivity under climate change in Uganda. Food Secur. 2025. https://doi.org/10.1007/s12571-025-01564-2.

21. Chitsiko RJ, Mutanga O, Dube T, Kutywayo D. Review of current models and approaches used for maize crop yield forecasting in sub-Saharan Africa and their potential use in early warning systems. Phys Chem Earth. 2022;127:103199. https://doi.org/10.1016/j.pce.2022.103199.
22. Joshi A, et al. An explainable Bi-LSTM model for winter wheat yield prediction. Front Plant Sci. 2024;15:1–17. https://doi.org/10.3389/fpls.2024.1491493.
23. Wang Y, Feng K, Sun L, Xie Y, Song XP. Satellite-based soybean yield prediction in Argentina: a comparison between panel regression and deep learning methods. Comput Electron Agric. 2024. https://doi.org/10.1016/j.compag.2024.108978.
24. Jarray N. Machine learning for food security: current status, challenges, and future perspectives. Artif Intell Rev. 2023;56:3853.
25. Meroni M, Waldner F, Seguini L, Rembold F. Agricultural and forest meteorology yield forecasting with machine learning and small data: What gains for grains ? Agric For Meteorol. 2021;309:108555. https://doi.org/10.1016/j.agrformet.2021.108555.
26. Islam M. Explainable machine learning for efficient diabetes prediction using hyperparameter tuning, SHAP analysis, partial dependency, and LIME. Eng Rep. 2024. https://doi.org/10.1002/eng2.13080.
27. Bin S, Esichaikul V, Lameesa A, Forruque S, Gandomi AH. Results in engineering an approach for crop recommendation with uncertainty quantification based on machine learning for sustainable agricultural decision-making. Results Eng. 2025;26:105505. https://doi.org/10.1016/j.rineng.2025.105505.
28. Shams MY, Gamel SA, Talaat FM. Enhancing crop recommendation systems with explainable artificial intelligence : a study on agricultural decision-making. Neural Comput & Appl. 2024;36(11):5695–714. https://doi.org/10.1007/s00521-023-09391-2.
29. Ikpe E, Omede UD, Alhassan YJ. Relationship between rainfall variability and maize yield in Apa local Government Area of Benue State, Nigeria. Gujarat J Ext Educ. 2024;37(2):7–13. https://doi.org/10.56572/gjoee.2024.37.2.0002.
30. Chagwiza C, Mapfumo P, Antwi M. Impact of climate change on maize yield. J Agribus Rural Dev. 2020;58(4):359.
31. Miao L, Zou Y, Cui X, Kattel GR, Shang Y, Zhu J. Predicting China's maize yield using multi-source datasets and machine learning algorithms. Remote Sens. 2024. https://doi.org/10.3390/rs16132417.
32. Díaz-Ramírez J. Machine learning and deep learning. Ingeniare. 2021;29(2):182–3. https://doi.org/10.4067/S0718-33052021000200180.
33. Panigrahi B, Kathala KCR, Sujatha M. A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models. Procedia Comput Sci. 2022;218(2022):2684–93. https://doi.org/10.1016/j.procs.2023.01.241.
34. Ge H, Zhang Q, Shen M, Qin Y, Wang L, Yuan C. Enhancing yield prediction in maize breeding using UAV-derived RGB imagery: a novel classification-integrated regression approach. Front Plant Sci. 2025;16(March):1–13. https://doi.org/10.3389/fpls.2025.1511871.
35. Shahhosseini M, Hu G, Archontoulis SV. Forecasting corn yield with machine learning ensembles. Front Plant Sci. 2020;11(July):1–16. https://doi.org/10.3389/fpls.2020.01120.
36. Jarray N, Ben Abbes A, Farah IR. A machine learning framework for cereal yield forecasting using heterogeneous data BT - intelligent systems design and applications. In: Abraham A, Pllana S, Casalino G, Ma K, Bajaj A, editors. Switzerland: Springer; 2023. pp. 21–30.
37. Filippi P, Han SY, Bishop TFA. On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies. Precis Agric. 2025;26:8.
38. Farag M, Emam A, Leonhardt J, Roscher R. Enhancing decision support in crop production : analyzing conformal prediction for uncertainty quantification. Comput Electron Agric. 2025;237(PB):110559. https://doi.org/10.1016/j.compag.2025.110559.
39. Bouni M, Hssina B, Douzi K, Douzi S. Interpretable machine learning techniques for an advanced crop recommendation model. J Electr Comput Eng. 2024. https://doi.org/10.1155/2024/7405217.
40. Shastri S, Kumar S, Mansotra V, Salgotra R. Advancing crop recommendation system with supervised machine learning and explainable artificial intelligence. Sci Rep. 2025;15(1):1–16. https://doi.org/10.1038/s41598-025-07003-8.
41. UBOS: Annual agricultural survey. In: Report, no. 2, pp. 2–5, 2022, [Online]. Available: https://www.ubos.org/wp-content/uploads/publications/07_2023AAS_2020_REPORT.pdf
42. Engen M, Sandø E, Sjølander BLO, Arenberg S, Gupta R, Goodwin M. Farm-scale crop yield prediction from multi-temporal data using deep hybrid neural networks. Agronomy. 2021;11(12):1–31. https://doi.org/10.3390/agronomy11122576.
43. Olofintuyi SS, Olajubu EA, Olanike D. An ensemble deep learning approach for predicting cocoa yield. Heliyon. 2023;9(4):e15245. https://doi.org/10.1016/j.heliyon.2023.e15245.
44. Alves F, et al. Bayesian ridge regression shows the best fit for SSR markers in *Psidium guajava* among Bayesian models. Sci Rep. 2021;11(1):0123456789. https://doi.org/10.1038/s41598-021-93120-z.
45. Mohale VZ, Obagbuwa IC. Evaluating machine learning-based intrusion detection systems with explainable AI: enhancing transparency and interpretability. Front Comput Sci. 2025;7:1–23. https://doi.org/10.3389/fcomp.2025.1520741.
46. Kalasampath K, Spoorthi KN, Sajeev S, Kuppa SS, Ajay K, Maruthamuthu A. A literature review on applications of explainable artificial intelligence (XAI). IEEE Access. 2025;13:41111–40. https://doi.org/10.1109/ACCESS.2025.3546681.
47. Akkem Y, Biswas SK, Varanasi A. Role of Explainable AI in crop recommendation technique of smart farming. Int J Intell Syst Appl. 2025;17(1):31–52. https://doi.org/10.5815/ijisa.2025.01.03.
48. Steurer M, Hill RJ, Pfeifer N, Hill RJ, Pfeifer N. Metrics for evaluating the performance of machine learning based automated valuation models based automated valuation models. J Prop Res. 2021;38(2):99–129. https://doi.org/10.1080/09599916.2020.1858937.
49. Jadon A, Patil A, Jadon S. A comprehensive survey of regression based loss functions for time series forecasting.
50. Tatachar AV. Comparative assessment of regression models based on model evaluation metrics. 2021; pp. 853–860.
51. Oikonomidis A, Catal C. Hybrid deep learning-based models for crop yield prediction. Appl Artif Intell. 2022. https://doi.org/10.1080/08839514.2022.2031823.
52. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast. 2006;22(4):679–88. https://doi.org/10.1016/j.ijforecast.2006.03.001.
53. Kim S, Kim H. A new metric of absolute percentage error for intermittent demand forecasts. Int J Forecast. 2016;32(3):669–79. https://doi.org/10.1016/j.ijforecast.2015.12.003.

54. Franses PH. A note on the Mean Absolute Scaled Error. Int J Forecast. 2016;32(1):20–2. https://doi.org/10.1016/j.ijforecast.2015.03.008.
55. Lionel BM, Musabe R, Gatera O, Twizere C. A comparative study of machine learning models in predicting crop yield. Discov Agric. 2025;3(1):151. https://doi.org/10.1007/S44279-025-00335-Z.
56. Nur L, Matsui T, Tanaka TST. Critical evaluation of the effects of a cross-validation strategy and machine learning optimization on the prediction accuracy and transferability of a soybean yield prediction model using UAV-based remote sensing. J Agric Food Res. 2024;16:101096. https://doi.org/10.1016/j.jafr.2024.101096.
57. Hoerl AE, Kennard RW. Ridge regression: applications to nonorthogonal problems. Technometrics. 1970;12(1):69–82. https://doi.org/10.1080/00401706.1970.10488635;WGROUP:STRING:PUBLICATION.
58. Chen T, Guestrin C. XGBoost : a scalable tree boosting system. 2016.
59. Kenduiywo BK, Miller S. Seasonal maize yield forecasting in South and East African countries using hybrid Earth observation models. Heliyon. 2024;10(13):e33449. https://doi.org/10.1016/j.heliyon.2024.e33449.
60. Lei J, G'Sell M, Rinaldo A, Tibshirani RJ, Wasserman L. "Distribution-free predictive inference for regression,." J Am Stat Assoc. 2018;113(523):1094–111. https://doi.org/10.1080/01621459.2017.1307116.
61. Romano Y, Patterson E, Candès EJ. Conformalized quantile regression. Adv Neural Inf Process Syst. 2019;32.
62. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;2017:4766–75.
63. Ribeiro MT, Singh S, Guestrin C. Why should I trust you?' Explaining the predictions of any classifier. In: Proceedings of ACM SIGKDD International Conference Knowledge Discover Data Min., vol. 13–17;2016. pp. 1135–1144. https://doi.org/10.1145/2939672.2939778/SUPPL_FILE/KDD2016_RIBEIRO_ANY_CLASSIFIER_01-ACM.MP4.

Publisher's Note