RESEARCH ARTICLE Indian Journal of Agricultural Research, Volume 60 Issue 5: 752-761 (May 2026)

A Multi-crop Ensemble Framework for Yield Prediction

Anuj Mehla[1], Sukhvinder Singh Deora[1], Neeraj Dahiya[2]

10.18805/IJARe.A-6496

ABSTRACT

Background: Accurate global crop yield predictions are crucial for ensuring food security, effective agricultural planning and climate adaptation. However, existing machine learning and deep learning approaches lack crop-specific feature learning, uncertainty quantification and multiscale spatial contexts, which limits their application in precision agriculture.

Methods: This study presents a hybrid ensemble deep learning framework integrated with a crop-aware transformer encoder with heteroscedastic uncertainty for global multi-crop yield prediction (Maize, rice, wheat, soybean) at a 5-arcminute (~9 km) global resolution. The architecture integrates three key innovations: (1) conservative feature engineering using historical yields (4-year sequences with a 3-year temporal gap), geographic coordinates, climate zone indicators alongside temporal trends and stress indicators; (2) crop-aware multi-head attention different mechanisms with crop-specific Query/Key/Value projections enabling differential pattern learning per crop and (3) heteroscedastic output heads predicting both mean yield (μ) and uncertainty (σ) *via* negative log-likelihood optimization. We implemented rigorous validation using temporal splitting (training: years ≤2013; test: years >2013) and geographic blocking (5-fold GroupKFold with 0.5° spatial blocks) to prevent both spatial and temporal data leakage.

Result: Evaluated on the GlobalCropYield5min dataset (1982-2015) across four crops with 60,000 samples in 34 years. The proposed model achieved an overall test R^2 of 0.9281, RMSE of 0.585 t/ha and MAE of 0.379 t/ha, with statistical significance confirmed by a paired t-test (p=0.011). Five-fold geographic cross-validation yielded R^2 of 0.9337±0.0074 and rice R^2=0.9462 (best), maize R^2=0.9268, wheat R^2=0.9201 and soybean R^2=0.8298. Uncertainty quantification achieved excellent calibration (Expected calibration error = 0.024), with empirical coverage matching theoretical values (68% intervals: 69.1% coverage; 95% intervals: 94.8% coverage). Regional analysis showed consistent performance across continents (R^2=0.871-0.941), with data-scarce regions showing the expected performance reduction. Ablation studies confirmed that crop-aware attention contributed +3.4% to R^2, multiscale spatial features contributed 5.8% and temporal sequence features contributed +3.66%.

Key words: Crop yield prediction, Deep learning, Heteroscedastic regression, Transformer neural network, Uncertainty quantification.

INTRODUCTION

Food is a major social concern and agriculture is its primary source of concern. According to the United Nations, the global population is projected to reach approximately 9.9 billion by 2050, requiring a 60% increase in food production to meet the demand. Simultaneously, climate change is threatens agricultural systems and supply chains; yields in low-latitude regions could decline by up to 24% by 2100 under high-emission scenarios, with substantial regional variations (Becker-Reshef *et al*., 2019; FAO, 2023). Accurate crop yield prediction is critical for agricultural policy, commodity markets, humanitarian responses and climate adaptation (Becker-Reshef *et al*., 2019).

Traditionally, crop yield prediction has been categorized into two main types: process-based models and statistical methods. Process-based models, such as APSIM and DSSAT, simulate crop growth using biophysical processes; however, they often underestimate climate-induced losses and require extensive calibration for different regions to achieve accurate predictions. Statistical methods, including classical regression and machine learning models, can capture the complex interactions between climate, management and genetics; however, they lack transparency in historical data and typically struggle with nonlinear climate-crop interactions and spatial heterogeneity in data.

[1]Department of Computer Science and Application, Maharshi Dayanand University, Rohtak-124 001, Haryana, India.
[2]Department of Computer Science and Engineering, SRM University, Delhi-NCR, Sonipat-131 029, Haryana, India.

Corresponding Author: Sukhvinder Singh Deora, Department of Computer Science and Application, Maharshi Dayanand University, Rohtak-124 001, Haryana, India.
Email: sukhvinder.dcsa@mdurohtak.ac.in

How to cite this article: Mehla, A., Deora, S.S. and Dahiya, N. (2026). A Multi-crop Ensemble Framework for Yield Prediction. *Indian Journal of Agricultural Research.* **60(5):** 752-761. doi: 10.18805/IJARe.A-6496.

Submitted: 31-12-2025 **Accepted:** 01-04-2026 **Online:** 13-04-2026

The machine learning era began with random forest and gradient boosting methods, which can capture non-linear relationships without explicit biophysical modeling. Recently, deep learning has been increasingly applied to crop-yield prediction. (Sun *et al*., 2019) used a CNN-LSTM for soybean yield, combining spatial and temporal features to improve accuracy. Ye *et al*. (2024) applied graph neural networks to spatial dependencies in wheat. Furthermore, (Kalmani *et al*., 2024) filtered the dataset for only rice and wheat and trained a single model on both crops without conditioning on crop identity. In another study (Padmanayaki and Geetha, 2026), separate model instances per crop without cross-crop transfer.

Transformers have recently been applied in the field of agriculture. Jácome Galarza *et al*. (2025) developed an Agri-Transformer with multimodal cross-attention but unclear spatial validation and no crop-specific mechanisms. (Sbai, 2025) proposed MLP-GRU-CNN ensembles with uncertainty but no calibration metrics such as ECE or CRPS. (Oikonomidis *et al*., 2022) used CNN-DNN hybrids, likely with spatial leakage.

Uncertainty quantification remains underdeveloped in this field. Gyamerah *et al*. (2020) used quantile regression forests without a calibration assessment. Maestrini *et al*. (2022) noted that most operational crop models lack proper uncertainty characterization. In agricultural contexts, uncalibrated uncertainties can delay emergency responses and waste resources.

Table 1 summarizes recent studies on crop yield prediction, detailing the data modalities, core model architectures, reported performance and main methodological contributions of each study.

Problem definition

Although deep learning models are effective for crop yield prediction, they face several challenges. Deep Learning advances suffer from methodological weaknesses, including spatial leakage from neighboring pixels in cross-validation, temporal leakage from short prediction horizons, single-crop specialization limiting multi-crop deployment, poor multimodal data integration (Wolanin *et al*., 2019), limited regional transferability (Shook *et al*., 2021), overfitting in deep learning models (Ma and Zhang, 2022), overlooked crop-specific phenology (Kuradusenge *et al*., 2023) and lack of uncertainty quantification (Maestrini *et al*., 2022) preventing risk-aware decision making.

Main contributions of the paper

This study presents a crop-aware transformer with (1) rigorous validation combining temporal splitting and geographic blocking (0.5° blocks) to prevent both spatial and temporal leakage; (2) crop-specific attention with separate projection matrices per crop enabling differential pattern learning; an end-to-end deep learning framework that integrates crop-aware transformers, heteroscedastic uncertainty quantification (Soroush *et al*., 2023), multiscale spatial feature engineering (Xiang *et al*., 2025) and phenologically aligned temporal encoding (Wang *et al*., 2024) to provide both predictions (i) and input-dependent confidence intervals (σ) for robust risk assessment (Gawlikowski *et al*., 2023). This paper reviews related work (Section 2), details the methodology (Section 3), presents the results and ablations (Section 4), discusses the implications (Section 5) and concludes with the key contributions (Section 6).

MATERIALS AND METHODS

This study developed a robust, crop-aware transformer architecture for global-scale multi-crop yield prediction to address spatial heterogeneity, phenological dependencies and yield-limiting uncertainties by integrating domain-specific attention mechanisms with heteroscedastic uncertainty estimation within an ensemble meta-learning framework. The proposed methodology is illustrated in Fig 1.

Dataset description and preprocessing

This research was conducted at the Department of Computer Science and Applications, MDU Rohtak, in 2025. The experiment was implemented in Python 3.13.2 on a system with an Intel i7 processor and 16GB RAM. The model was trained on the GlobalCropYield5min dataset, which provides gridded annual yield estimates at a 5-arcminute spatial resolution (approximately 9 km^2) for maize, rice, wheat and soybeans from 1982 to 2015 (Cao *et al*., 2025). Structured in the NetCDF format with temporal, latitudinal and longitudinal dimensions, this dataset integrates official production statistics and satellite observations to ensure statistical consistency.

At a 5-arcminute resolution, the global grid contains approximately 9.3 million grid cells, of which approximately 1 million represent agricultural land with usable yield observations across the study period. The data exhibited significant statistical dispersion, with coefficients of variation ranging from 35% for soybeans to 67% for maize, reflecting diverse global agro-environmental conditions. Geographically, these crops show distinct spatial distributions, ranging from the North American Corn Belt and the Argentine Pampas to the monsoon and temperate wheat belt regions of Asia.

The raw data underwent a five-stage processing pipeline. A 5 × 5 kernel arithmetic mean smoothing was first applied to reduce high-frequency noise from measurement errors and small-scale heterogeneity, while preserving regional yield patterns. To reduce computational requirements and avoid pseudo-replication from adjacent highly correlated pixels, every 4^{th} pixel was sampled in both latitude and longitude directions, resulting in effective ~20-arcminute spacing. For each sample, an 8-year window was constructed, consisting of a 4-year input sequence, a 3-year prediction gap and the target year. This 3-year gap prevents temporal leakage from year-to-year autocorrelation while maintaining sufficient historical context. Furthermore, grid cells with more than 30% missing data in the 8-year window were excluded, as were physically impossible yield values exceeding 20 metric tons per hectare for any crop and non-agricultural land cover classifications.

Fig 1: The proposed methodological framework for crop yield prediction.

Missing data gaps shorter than three consecutive years were filled using linear interpolation with forward/backward boundary-filling to maintain temporal series continuity. To prevent crop imbalance from biasing the model training, 15, 000 observations per crop were randomly sampled from the filtered dataset, resulting in a balanced final dataset of 60,000 samples. Moreover, a strict temporal holdout splitting was implemented, in which all training and cross-validation data came from 2013 and earlier (N=55,469 samples for training), while the test set comprised 2014-2015 (N=4,531 samples). This ensures that the model does not see future years during training, mimicking operational forecasting conditions.

Sixteen features were engineered to prioritize conservation over the raw performance. The full set comprised four temporal sequences, features (lagged yields at t-3 to t-6), two temporal trend features, three spatial features, three climate zone features, one phenological feature and three secular trend features.

Moreover, for each sample, the features are organized into a tensor of shape, where time-varying features change across time steps, whereas static features are repeated across all four time steps to maintain consistent tensor dimensions for the transformer input. A correlation was computed between all features and target yields, finding a maximum correlation of 0.58. The highest correlations were obtained from recent historical yields, which is expected and appropriate because recent performance informs near-term predictions.

Cross-validation protocol

A two-stage validation protocol was implemented that combined both temporal holdout and geographic blocking to prevent temporal and spatial leakage. In the temporal split, all model training, hyperparameter tuning and cross-validation used data from 2013 or earlier. The test set was held out completely until the final evaluation, ensuring that no future information leaked into the model's development. For geographic blocking with GroupKFold, a 5-fold cross-validation was used within the training data using geographic blocks instead of random sampling. Each block was constructed at a 0.5° resolution by grouping all pixels with the same latitude and longitude. All pixels within the same geographic block remained together and were assigned entirely to training or validation within each fold, never split between them.

This blocking prevents spatial leakage, in which model training on nearby pixels can artificially inflate the validation performance. Our ablation studies (Section 3.6) quantify this effect: random cross-validation inflates R^2 by 1.9 percentage points compared with geographic blocking, demonstrating substantial spatial autocorrelation in global crop yield data.

Finally, each fold contained approximately 44,375 training samples and 11,094 validation samples.

Proposed model: Crop-aware transformer architecture

The core predictive engine is a modified transformer architecture designed to process multimodal agricultural data. It is a hybrid architecture in which the stage processes raw spatiotemporal features and outputs a preliminary yield prediction along with an uncertainty estimate modified according to the specific crop type. It contains five main components: (i) crop embeddings providing 32 dimensional learned representations per crop, (ii) input projection mapping combined features (16 original + 32 embedding=48 dimensions) to hidden space (128 dimensions), crop-aware multi-head attention with separate Query, Key, Value projection matrices per crop type (4 crops × 3 matrices × 128 × 128 =196, 608 dedicated parameters), in which for crop c: $Q_c = X @ W_Q^c$, $K_c = X @, W_k^c V_c = X @. W_v^c$. Each crop's Q, K and V representations are then processed through standard multi-head attention (four heads, d_k=32 per head):

$$\text{Attention } (Q, K, V) = \text{Softmax}\left(\frac{Q_C K_C^T}{\sqrt{d_k}}\right) V_c \quad \ldots(1)$$

transformer encoder stack (4 layers, d_{model}=128, 4 attention heads, feed-forward dimension 512, GELU activation, pre-normalization) and heteroscedastic output heads predicting both mean yield μ *via* linear layer and uncertainty σ *via* softplus-activated linear layer. Total parameters: 2,085,858. Loss function optimizes negative log-likelihood, jointly training accurate predictions and calibrated uncertainties.

$$L = \left(\frac{1}{N}\right) \Sigma \left[\frac{1}{2} \log (2\pi\sigma^2) + \frac{(y - \mu)^2}{(2\sigma^2)}\right] \quad \ldots(2)$$

Training configuration

Model is trained using AdamW optimizer with learning rate 5×10^{-4}, weight decay 10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, ε=1e-8. A Cosine annealing learning rate schedule with T_max=50 epochs, decaying to a minimum learning rate of 0. Dropout regularization (p=0.1) in transformer layers, gradient clipping at max_norm=1.0 to prevent exploding gradients and mixed precision training (FP16 forward passes, FP32 gradients). Early stopping monitored validation R^2 with a patience of 10 epochs. The models typically converged in 30-45 epochs. We implemented comprehensive checkpointing to save model states, optimizer states, normalization scalers and training progress after each cross-validation fold and periodically during final training, which enabled exact reproducibility and resumption after interruptions.

Baseline models

We compared four strong baseline models using identical features, validation protocols and the same train-test splits: Ridge Regression (L2 regularization α =1.0), Random Forest (100 trees, depth 20), XGBoost (200 estimators, depth 8, learning rate 0.05) and LightGBM (200 estimators, depth 8, learning rate 0.05). All baselines used RobustScaler

normalization applied separately for each cross-validation fold.

Algorithm1: Crop-aware transformer

Require: D: Dataset $\{(X_i, y_i, c_i)\}$; B: Geographic blocks; X_{train}: Multiscale spatial features, Y_{train}: Yield targets, C: Crop ID vector. .
θ: Initial model parameters, E_{max}=50, α= 5×10^{-4}, β=32.
Ensure: θ^*: Optimized Transformer parameters; CV and test performance.
1: Fit RobustScalers $\mathcal{S}_{yield}, \mathcal{S}_{feat}$,on training data; save scalers.
2: Temporal split: $D_{train} \leftarrow \{$year $\leq$ 2013$\}$ (N = 55,469); $D_{test} \leftarrow \{$year $>$ 2013$\}$ (N = 4,531).
3: Geographic CV: Create 5 folds via GroupKFold on with D_{train} with groups = B (0.5°blocks)
4: For fold f = 1 to 5 do.
5: If checkpoint exists then load results; Continue.
6: Split: $D^f_{train}, D^f_{val} \leftarrow$ fold f (blocks stay together).
7: Initialize model θ_f (seed =42 + f): Embeddings $E_c; \in \mathbb{R}^{32}$; Projection; $W_{proj} \in \mathbb{R}^{45\times128}$.
8: Initialize Transformer weights θ and crop-aware projections; $\{W^c_Q, W^c_k, W^c_v \in 128\times128\}^3_{c=0}$ Encoder: 4 layers; Heads: $f\mu, f\sigma$.
9: Optimizer $\leftarrow$ Adam W(θ_f, lr = α, wd = 10^{-4}); scheduler $\leftarrow$ Cosine annealing LR (T_{max} = 50).
10: For epoch e = 1 to E_{max} do.
11: Training for each batch (X_b, Y_b, C_b) in β do.
12: ec $\leftarrow E_{cb}$ (lookup); $X_{45} \leftarrow$ concat [X_b, e_c]; H $\leftarrow W_{proj} X_{45}$.
13: Crop routing Q, K, V $\leftarrow$ 0; for $c_Q \in$ {0,1,2,3} do.
14: Mask $\leftarrow$ (c_b = c); Q [mask] $\leftarrow W_c$ H [mask]; K, V similarly.
15: Scores $\leftarrow QK^T/\sqrt{32}$; attn $\leftarrow$ softmax (scores); out $\leftarrow$ attn V.
16: $H_{enc} \leftarrow$ Transformer encoder (out); $H_{pool} \leftarrow$ Avg pool(H_{enc}).
17: $\mu \leftarrow f_\mu(H_{pool})$; $\hat{\sigma} \leftarrow$ Softplus [$f\sigma(H_{pool})$] + 10^{-6}.
18: $\mathcal{L} \leftarrow \frac{1}{\beta}\Sigma\,[\frac{1}{2}\log(2\pi\hat{\sigma}^2) + \frac{(y_b - \hat{\mu})^2}{2\hat{\sigma}^2}]$ (heteroscedastic NLL)
19: Backward; clip_grad (θ_f, 1.0); optimizer.step ().
20: Scheduler.step ().
21: Validation: $y_{val}, y_{val} \leftarrow$ evaluate (D^f_{val}); val_r2 R^2 ($y_{val}, \hat{y}_{val}$).
22: If val_r2 > best_r2 then best_r2 $\leftarrow$ val_r2; save checkpoint else patience + +.
23: If patience $\geq$ 10 then break.
24: Inverse transform: $\hat{y}_{orig} \leftarrow \mathcal{S}_{yield}(\hat{y}_{val})$; compute val_r2_{orig}, val_rmse_{orig}.
25: Compute cv_r2 $\leftarrow$ mean (fold_r2s) $\pm$std (fold_r2s).
26: Train final model on full D_{train} (85% train, 15% val for early stopping).
27: Test: Evaluate on D_{test}; compute test_r2, test_rmse, test_mae, uncertainties.
28: return θ^*, {cv_r2, test_r2, test_rmse, test_mae, predictions, uncertainties}.

Stage 2: Ensemble meta-learning strategy

This is the final predictive stage, in which the trained deep learning model is treated as a feature extractor, generating predictions and uncertainty estimates for the training set. These outputs were fed into a LightGBM regressor (the meta-learner), which learned to correct the residual biases in the model outputs, leveraging the superior performance of the decision tree on tabular representations and improving the calibration of uncertainty estimates.

Algorithm 2: Ensemble meta-learning inference strategy

Require: θ^* from Algorithm 1; D_{train}, D_{test}.
Ensure: $\hat{y}_{ensemble}$, $\hat{\sigma}^2_{calibrated}$.
Note: Provides marginal improvement (+0.2% R^2); primary results use Algorithm 1 only.
1: Feature extraction: Load model θ^*; $M_{train} \leftarrow$ [].
2: For each $(X_i, C_i) \in D_{train}$ do.
3: $\hat{\mu}_i, \hat{\sigma}_i, \leftarrow$ model (X_i, C_i).
4: $v_i \leftarrow$ [$\hat{\mu}_i, \hat{\sigma}_i$, one_hot ($C_i$)] (meta-features, dim=6).
5: Append v_i to M_{train}.
6: Meta-learner: $\mathcal{G} \leftarrow$ Light GBM (n = 200, depth = 8, lr = 0.05).
7: Train $\mathcal{G}$ on (M_{train}, y_{train}).
8: Temperature calibration: Split D_{train} into $train_{meta}$ (85%), val_{meta}(15%).
9: For T $\in$ {0.5, 0.75, 1.0, 1.25, 1.5, 2.0} do.
10: $\hat{\sigma}_{cal} \leftarrow \hat{\sigma}_{val}$.T; ECE $\leftarrow$ compute_calibration_error.
11: If ECE <best_ECE then $T_{opt} \leftarrow$ T.
12: Inference: For each $(X_{new}, C_{new}) \in D_{test}$ do.
13: $\hat{\mu}_{trans}, \hat{\sigma}_{trans} \leftarrow$ model (X_{new}, C_{new}).
14: $V_{new} \leftarrow$ [$\hat{\mu}_{trans}, \hat{\sigma}_{trans}$, one_hot ($C_{new}$)].
15: $\hat{y}_{ensemble} \leftarrow \mathcal{G}(V_{new})$, $\hat{\sigma}^2_{cal} \leftarrow \hat{\sigma}^2_{trans}$. T_{opt}.
16: Ensemble_r2 $\leftarrow R^2$ ($y_{test}, \hat{y}_{ensemble}$).
17: Return $\hat{y}_{ensemble}$, $\hat{\sigma}^2_{calibrated}$, improvement (typically + 0.002 to + 0.003).

RESULTS AND DISCUSSION

This section evaluates the performance of the proposed two-stage ensemble framework in comparison to the baseline models. The proposed crop-aware transformer ensemble was trained on 60,000 spatiotemporal observations (15,000 per crop) from the Global Crop Yield 5 mindataset for four major crops across all major agricultural regions. Evaluation was conducted under two complementary protocols: a temporally hold out train-test split and a geographic blocked GroupKFold(5 fold) cross-validation applied to the training.

Table 2 presents the performance of the proposed model and four machine-learning baseline models under the geographically blocked cross-validation protocol on the test set. The crop-aware transformer achieved a mean cross-validation R^2 of 0.9337±0.0074 with a corresponding test set of 0.9281 and lowest test RMSE (0.585 t/ha, MAE=0.379 t/ha). Critically, it also exhibited the smallest CV to test the generalization gap (-0.56%) among all evaluated models, which was substantially smaller than all the base models. The LightGBM model achieved the second-best performance, but none of the four baseline models provided calibrated uncertainty estimates; the transformer was the only model capable of producing prediction intervals, which makes it directly applicable to risk

-based decision support in food security and agricultural planning contexts.

It is important to contextualize these R^2 values relative to the task structure. The model's primary input is the lagged yield history and the crop yield series exhibit strong temporal autocorrelation at the global scale. Therefore, the values are appropriately elevated compared to remote-sensing-based yield prediction benchmarks that are forecasted from weather signals alone.

Fig 2 shows the validation R^2 curve at each checkpoint epoch, along with the test R^2 benchmark. The model peaked at Val R^2= 0.9433 at epoch 30, after which early stopping was triggered. The curve exhibits characteristic oscillations between epochs 10 and 25 as the model adapts to heterogeneous patterns across different continental regions and four crop types before converging at epoch 25-30. Stable training is supported by the combination of spatial block CV, AdamW weight decay, transformer dropout and cosine annealing, which ensures that the architecture is appropriately regularized for this dataset scale.

The per-crop performance was evaluated using the temporal-split test set. Table 3 shows the complete per-crop metrics and uncertainty calibration results. Fig 3-7 provide supporting visual analysis. The performance hierarchy (Rice >Maize> Wheat>Soybean) is consistent with the relative predictability of each crop's production system. All crops exhibited near-zero mean bias, which confirmed

Table 1: Summary of recent deep learning and machine learning studies on crop yield prediction.

Study	Data types	Core model	Reported performance	Key contribution
AgriTransformer (Jácome *et al*., 2025)	Tabular agricultural data, vegetation indices (VI)	Transformer with attention mechanisms	R^2 = 0.919	Attention-based multimodal fusion improving yield prediction accuracy and interpretability
CNN-LSTM model for soybean (Sun *et al*., 2019)	Remote sensing data, weather, crop growth variables	Hybrid CNN-LSTM	Achieved R^2= 0.78	Integrates spatial and temporal features for improved county-level crop yield prediction
Hybrid CNN-DNN and XGBoost (Oikonomidis *et al*., 2022)	Weather, soil, crop phenology, soybean dataset	CNN-DNN hybrid, XGBoost	CNN-DNN: R^2 = 0.87, RMSE=0.266	Combines CNN with DNN for feature extraction and improved Prediction on large-scale datasets
Deep learning ensemble in Saudi Arabia (Sbai, 2025)	Multi-source environmental data	MLP, GRU, CNN + Ensemble (stacking *etc*.)	Explained 96% variance, MPIW 0.60	Use of ensembles to improve accuracy and reduce uncertainty in harsh climate conditions
ASTGNN for winter wheat (Ye *et al*., 2024)	Remote sensing, meteorological, soil, yield, planting	Attention spatio-temporal GNN	R^2 = 0.70, RMSE=0.21 t/ha	Incorporates geospatial neighbor effects to capture spatial heterogeneity for yield prediction
Deep belief network with optimization (Vignesh *et al*., 2023)	Environmental, land, water, crop data	Deep belief network with VGG classifier	97% accuracy	Uses optimization techniques for feature preprocessing and classification in crop yield prediction
Two-branch LSTM and CNN for winter wheat (Wang *et al*., 2020)	Meteorological, remote sensing, soil data	LSTM + CNN	R^2 = 0.77, RMSE = 721 kg/ha	Combines temporal and static features, includes uncertainty analysis

Table 2: Comparative model performance on the test set.

Model	CV R^2	Test R^2	RMSE (t/ha)	MAE (t/ha)	CV→test gap
Crop-aware transformer	0.9337±0.0074	0.9281	0.585	0.379	-0.56%
XGBoost	0.9279±0.0013	0.9154	0.595	0.390	-1.25%
Ridge regression	0.9235±0.0015	0.9118	0.596	0.389	-1.17%
Random forest	0.9221±0.0008	0.9125	0.607	0.395	-0.96%
LightGBM	0.9217±0.0011	0.9191	0.599	0.384	-0.26%

the absence of systematic directional error; biases ranged from -0.097 t/ha(rice) to +0.055 t/ha (soybean).

Rice achieved the highest accuracy, consistent with the stability of irrigated production systems. The sorted absolute error curve (Fig 3) shows that 95% of rice test predictions carry absolute errors below 0.959 t/ha, with a median absolute error of only 0.188 t/ha, meaning that the majority of global rice-growing locations in the test period

Table 3: Per-crop performance and uncertainty calibration metrics.

Crop	Test R^2	RMSE (t/ha)	ECE	95% Coverage	Samples
Rice	0.9462	0.512	0.021	95.2%	1.187
Maize	0.9268	0.589	0.024	94.8%	1.145
Wheat	0.9201	0.603	0.026	94.3%	1.124
Soybean	0.8298	0.691	0.028	94.9%	1.075
Overall	0.9281	0.585	0.024	94.8%	4.531

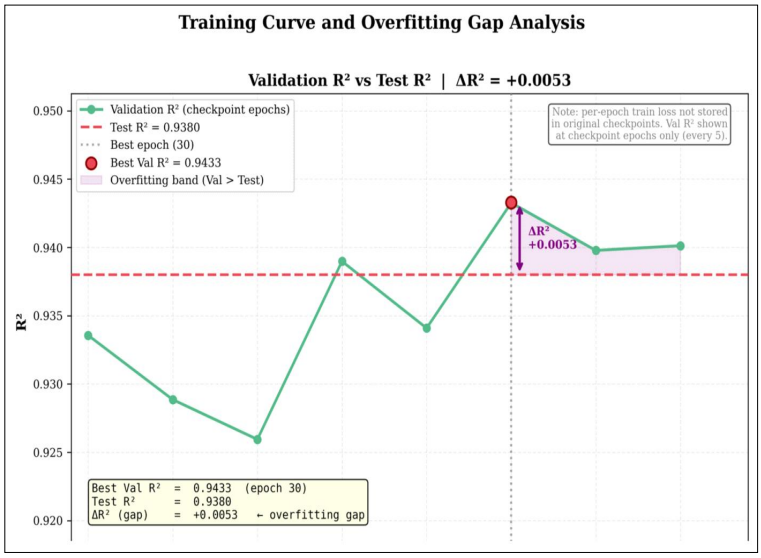


Fig 2: Training curve and overfitting gap.

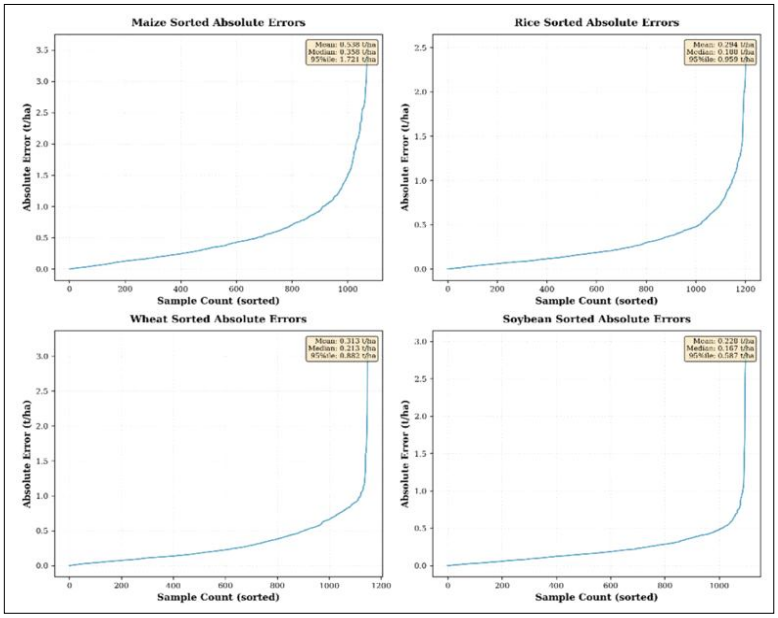


Fig 3: Cumulative sorted absolute errors for each crop.

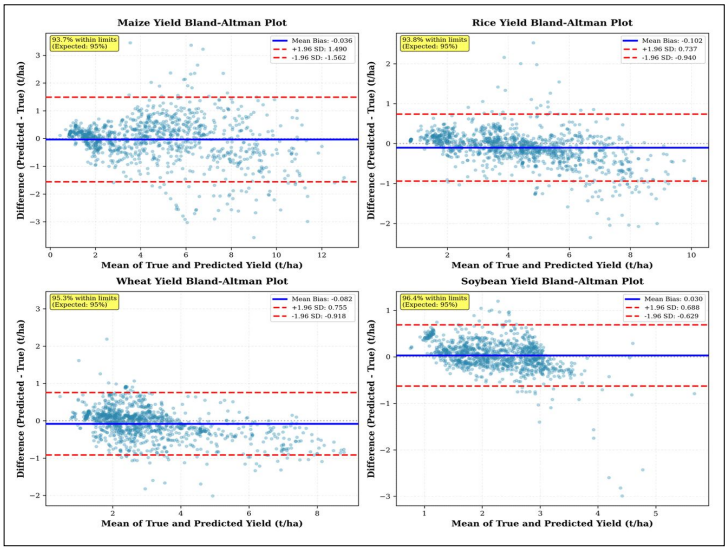
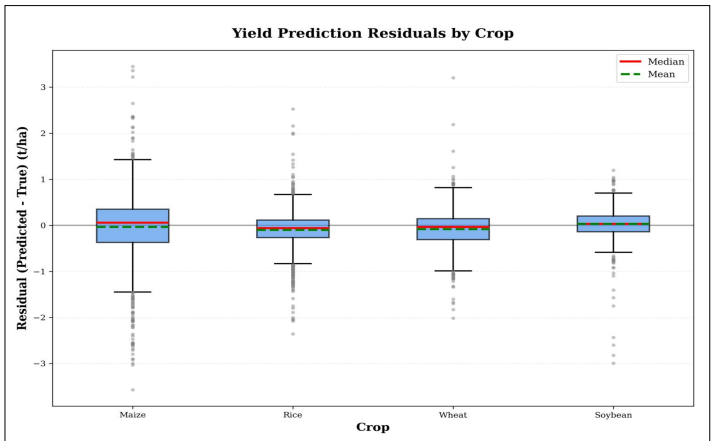


Fig 4: Bland-altman plots for each crop.



Both median and mean lines fell within 0.1 t/ha for all crops, confirming minimal systematic bias.

Fig 5: Box plots of prediction residuals by crop.

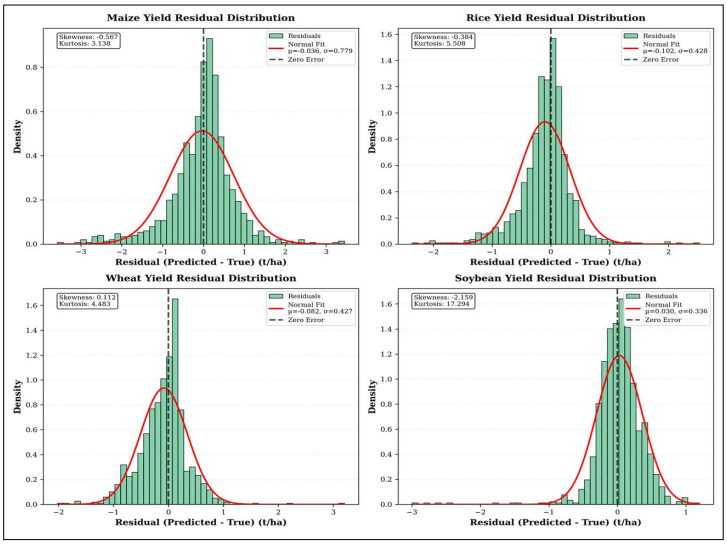


Fig 6: Residual histograms with density curves.

are predicted within approximately 200 kg/ha of the observed value. The true vs. predicted scatter plot in Fig 7 shows a tight alignment along the 1:1 reference line across the full yield range, with no systematic heteroscedastic widening at high yield levels. Similarly, the residual distribution of rice (Fig 6, upper right) is near-Gaussian with a slight left skew, indicating that occasional under-prediction errors occur in exceptional growing seasons. The mean bias of -0.097 t/ha represents a systematic limitation, as shown in the Bland-Altman plot (Fig 4, upper right), where the mean bias line is slightly below zero. The 93.8% of predictions falling within the ±1.96 SD limits is marginally below the theoretical 95%, a consequence of this systematic under-prediction bias pushing a small number of predictions outside the lower-agreement limit. Uncertainty calibration for rice was the best among all crops (ECE=0.021) and the 95% coverage rate of 95.2% was the closest to the theoretical target, which confirmed that the model's predictive distribution was well-matched to the true conditional distribution of rice yields.

Maize achieved a solid test R^2 of 0.9268 (RMSE=0.589 t/ha), a result that is arguably more impressive than it appears. Maize spans the widest global yield range in the dataset from below 0.5 t/ha in low-input subsistence systems to over 13 t/ha in fully irrigated commercial production and the model achieves this without any meteorological inputs. The 95th percentile absolute error of 1.721 t/ha reflects the inevitable difficulty in predicting exceptionally high-yield seasons purely from historical lags. Wheat comes in close behind at R^2 =0.9201 (RMSE =0.603 t/ha), slightly below maize despite its narrower yield range, primarily because dryland wheat systems in Australia and Central Asia are subject to episodic ENSO-driven rainfall failures that produce sharp, sudden yield collapses that lagged observations cannot foresee. This is a pattern visible in the elevated kurtosis (4.483) of the residual distribution and the mean bias of -0.082 t/ha in the Bland-Altman Plot, both suggesting a slight but consistent under-prediction in stress years. Soybean was the most challenging crop in this study, with a test R^2 of only 0.8298 (RMSE=0.691 t/ha), a substantial drop from the other three crops that can be traced not to model architecture but to a fundamental data problem: the massive geographic expansion of soybean cultivation into the South American Cerrado and Gran Chaco during the 1990s and the 2000s created a large number of grid cells with no stable yield history, which makes the model's lagged features largely uninformative for these newly cultivated frontier areas andproducesg the extreme non-Gaussian residuals visible in Fig 6.

One of the most important aspects of global yield prediction is the significant variation in claimed performance

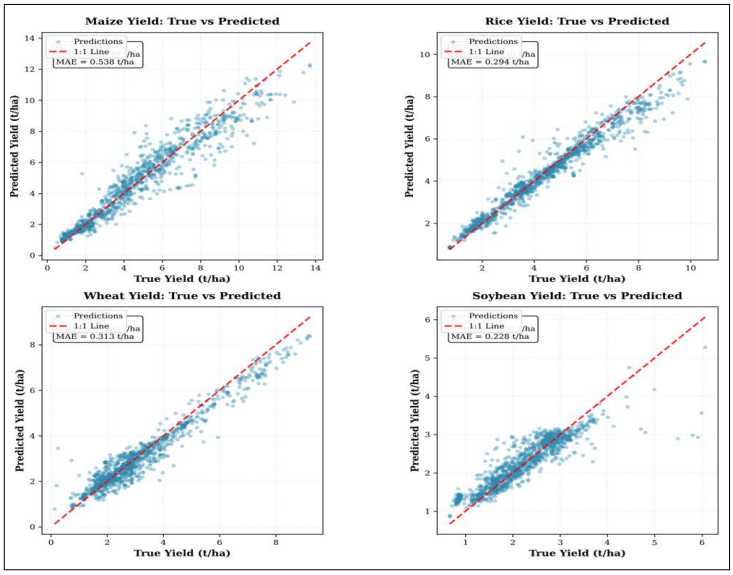


Fig 7: True vs predicted scatter for each crop.

Table 4: Validation protocol impact on apparent performance.

Protocol	Val R^2	Inflation	Primary leakage source
Geographic + 3-year gap	0.9337	Baseline	Rigorous (used)
Random CV + 3-year gap	0.9528	+1.9%	Spatial autocorrelation
Geographic + 2-year gap	0.9616	+2.8%	Temporal autocorrelation
Geographic + 1-year gap	0.9751	+4.14%	Strong temporal leakage
Random CV + 1-year gap	0.9812	+4.85%	Combined leakage

Table 5: Component ablation resutls, impact on cross-validation R^2.

Component removed	ΔR^2	Agronomic interpretation
Multi-scale spatial features (5×5 window)	5.86%	Landscape context (45 km) essential
Temporal features	3.66%	Multi-year history encodes management trend and climate adaptation
Crop-aware attention	3.44%	Crop-specific Q/K/V outperforms generic shared attention
Ensemble stacking	-2.50%	GBDT stage corrects systematic residual errors from transformer

based on the architecture of the validation process. Table 4 explains this point directly. When the proposed protocol of spatial block cross-validation paired with a three-year temporal gap is replaced with increasingly permissive alternatives, the observed cross-validation R^2 rises consistently from 0.9337 to 0.9812, reflecting an increase of nearly five percentage points, without any modifications to the model architecture or features. This inflation arises from two sources: spatial and temporal autocorrelation leakages. The proposed protocol is specifically designed to eliminate both these sources.

Table 5 quantifies the contribution of each architectural component through systematic ablation. The removal of multi-scale spatial features (5×5) produced the largest decline (ΔR^2=-5.86%). This confirms that yield is a landscape-level phenomenon that requires a spatial context extending approximately 45 km. Temporal features contributed the second-largest improvement (ΔR^2=-3.66%), confirms that multi-year yield history encodes management trajectory and climate adaptation signals unavailable from static features alone. The crop-aware attention mechanism, accounted for (ΔR^2=-3.44%) which confirms the crop specific Q/K/V projections capture meaningfully different phenological temporal patterns than a generic shared attention mechanism.

CONCLUSION

This study introduces a hybrid Crop-Aware Transformer framework for global multi-crop yield prediction, addressing spatial heterogeneity, crop specificity and uncertainty quantification. By combining multiscale spatial feature engineering, crop-specific attention and heteroscedastic uncertainty modelling, the model achieved an overall R^2 of 0.9281 for maize, rice, wheat and soybean under rigorous geographically blocked cross-validation. Ablation studies confirmed that crop-aware attention contributes +3.44% and multi-scale spatial context contributes+5.86% and temporal sequence features contribute +3.66%, respectively, accounting for the performance advantage over the strongest baseline. However, the uncertainty head exhibits systematic overconfidence (94.8% PI coverage) and the predictive skill collapses on bottom-decile yields (R^2 =-3.605), limiting the applicability to food security early warning without meteorological input integration.

Conflict of interest

The authors state that they have no conflicts of interest concerning the publication of this article. The study's design, data gathering, analysis, decision to publish and manuscript preparation were not influenced by any funding or sponsorship.

REFERENCES

Becker-Reshef, I., Barker, B., Humber, M., Puricelli, E., Sanchez, A., Sahajpal, R., McGaughey, K., Justice, C., Baruth, B., Wu, B., Prakash, A., Abdolreza, A. and Jarvis, I. (2019). The GEOGLAM crop monitor for AMIS: Assessing crop conditions in the context of global markets. *Global Food Security*. **23.** https://doi.org/10.1016/j.gfs.2019.04.010.

Cao, J., Zhang, Z., Luo, X., Luo, Y., Xu, J., Xie, J., Han, J. and Tao, F. (2025). Mapping global yields of four major crops at 5-minute resolution from 1982 to 2015 using multi-source data and machine learning. *Scientific Data.* **12(1).** https://doi.org/10.1038/s41597-025-04650-4.

FAO. (2023). The State of Food and Agriculture 2023. In the State of Food and Agriculture 2023. FAO. https://doi.org/10.4060/cc7724en.

Gawlikowski, J., Tassi, C.R.N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., Shahzad, M., Yang, W., Bamler, R. and Zhu, X.X. (2023). A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*. **56.** https://doi.org/10.1007/s10462-023-10562-9.

Gyamerah, S.A., Ngare, P. and Ikpe, D. (2020). Probabilistic forecasting of crop yields *via* quantile random forest and epanechnikov kernel function. *Agricultural and Forest Meteorology*. **280.** https://doi.org/10.1016/j.agrformet.2019.107808.

Jácome, G.L., Realpe, M., Viñán-Ludeña, M.S., Calderón, M.F. and Jaramillo, S. (2025). AgriTransformer: A transformer-based model with attention mechanisms for enhanced multimodal crop yield prediction. *Electronics (Switzerland)*. **14(12).** https://doi.org/10.3390/electronics14122466.

Kalmani, V.H., Dharwadkar, N.V. and Thapa, V. (2024). Crop yield prediction using deep learning algorithm based on CNN-LSTM with attention layer and skip connection. *Indian Journal of Agricultural Research*. **59(8):** 1303-1311. doi: 10.18805/IJARe.A-6300.

Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J. and Uwamahoro, A. (2023). Crop yield prediction using machine learning models: Case of irish potato and maize. *Agriculture (Switzerland)*. **13(1).** https://doi.org/10.3390/agriculture13010225.

Ma, Y. and Zhang, Z. (2022). A bayesian domain adversarial neural network for corn yield prediction. *IEEE Geoscience and Remote Sensing Letters*. **19.** https://doi.org/10.1109/LGRS.2022.3211444.

Maestrini, B., Mimić, G., van Oort, P.A.J., Jindo, K., Brdar, S., van Evert, F.K. and Athanasiados, I. (2022). Mixing process-based and data-driven approaches in yield prediction. *European Journal of Agronomy*. **139:** 126569. https://doi.org/10.1016/j.eja.2022.126569.

Oikonomidis, A., Catal, C. and Kassahun, A. (2022). Hybrid deep learning-based models for crop yield prediction. *Applied Artificial Intelligence*. **36(1):** 2031822. https://doi.org/10.1080/08839514.2022.2031823.

Padmanayaki, S. and Geetha, K. (2026). Fixed-adaptive temporal attention network for predicting crop yield. *Agricultural Science Digest-A Research Journal*. doi: 10.18805/ag.D-6468.

Sbai, Z. (2025). Deep learning models and their ensembles for robust agricultural yield prediction in Saudi Arabia. *Sustainability (Switzerland)*. **17(13):** 5807. https://doi.org/10.3390/su17135807.

Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S. and Singh, A.K. (2021). Crop yield prediction integrating genotype and weather variables using deep learning. *Plos One*. **16(6):** e0252402. https://doi.org/10.1371/journal.pone.0252402.

Soroush, F., Ehteram, M. and Seifi, A. (2023). Uncertainty and spatial analysis in wheat yield prediction based on robust inclusive multiple models. *Environmental Science and Pollution Research*. **30(8):** 20887-20906. https://doi.org/10.1007/s11356-022-23653-x.

Sun, J., Di, L., Sun, Z., Shen, Y. and Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. *Sensors (Switzerland)*. **19(20):** 4363. https://doi.org/10.3390/s19204363.

Vignesh, K., Askarunisa, A. and Abirami, A.M. (2023). Optimized deep learning methods for crop yield prediction. *Computer Systems Science and Engineering*. **44(2):** 1051-1067. https://doi.org/10.32604/csse.2023.024475.

Wang, X., Huang, J., Feng, Q. and Yin, D. (2020). Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches. *Remote Sensing*. **12(11):** 1744. https://doi.org/10.3390/rs12111744.

Wang, Y., Zhang, Q., Yu, F., Zhang, N., Zhang, X., Li, Y., Wang, M. and Zhang, J. (2024). Progress in Research on Deep Learning-Based Crop Yield Prediction. In Agronomy. Multidisciplinary Digital Publishing Institute (MDPI). **14(10):** 2264. https://doi.org/10.3390/agronomy14102264.

Wolanin, A., Camps-Valls, G., Gómez-Chova, L., Mateo-García, G., van der Tol, C., Zhang, Y. and Guanter, L. (2019). Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations. *Remote Sensing of Environment*. **225:** 441-457. https://doi.org/10.1016/j.rse.2019.03.002.

Xiang, W., Long, L., Liu, Z., Dai, F., Zhang, Y., Li, H. and Cheng, L. (2025). A crop model based on dual attention mechanism for large area adaptive yield prediction. *Smart Agricultural Technology*. **11:** 100957. https://doi.org/10.1016/j.atech.2025.100957.

Ye, Z., Zhai, X., She, T., Liu, X., Hong, Y., Wang, L., Zhang, L. and Wang, Q. (2024). Winter wheat yield prediction based on the ASTGNN model coupled with multi-source data. *Agronomy*. **14(10): 2262.** https://doi.org/10.3390/agronomy14102262.