Research Square

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

Yield-Graph: Multi-stage Growth-aware Maize Yield Prediction via Graph Neural Networks

Jiahui Wang

wangjiahui2928@163.com

Beijing University of Technology https://orcid.org/0009-0004-8398-9869

Yong Zhang

Beijing University of Technology

Bo Li

University of Macau

Yuqing Zhang

Beijing University of Technology

Xinglin Piao

Beijing University of Technology

Aiwen Wang

Beijing University of Technology

Xiangyu Zhao

Beijing Academy of Agriculture and Forestry Sciences

Kaiyi Wang

Beijing Academy of Agriculture and Forestry Sciences https://orcid.org/0009-0000-5365-7178

Research Article

Keywords: maize yield prediction, multi-stage maize growth data, data imputation, hypergraph attention network, bipartite graph neural network

Posted Date: September 26th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7532868/v1

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

Yield-Graph: Multi-stage Growth-aware Maize Yield Prediction via Graph Neural Networks

Jiahui Wang[1], Yong Zhang[1], Bo Li[2], Yuqing Zhang[1], Xinglin Piao[1], Aiwen Wang[1], Xiangyu Zhao[3], Kaiyi Wang[3]*

[1]Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing Institute of Artificial Intelligence, School of Information Science and Technology, Beijing University of Technology, Beijing, 100124, China

[2]PAMI Research Group, Department of Computer and Information Science, University of Macau, Taipa, Macau SAR, 999078, China

[3]Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, 100097, China

Abstract

Accurate yield prediction before maize harvest is crucial for advancing agricultural management and ensuring food security. Unlike conventional approaches that rely on phenotypes from a single growth stage, this study models multiple traits across different developmental stages, all targeting final yield, thereby uncovering their dynamic and cumulative contributions. We introduce Yield-Graph, an innovative framework that integrates multi-stage phenotypic data for yield prediction. The method employs a bipartite graph structure to impute missing trait values at each stage and leverages a hypergraph attention mechanism to capture high-order sample relationships. Comprehensive benchmark experiments demonstrate that Yield-Graph consistently outperforms traditional machine learning and graph-based models in both trait completion and yield prediction. Moreover, the framework exhibits strong robustness across growth stages, high adaptability to regional variations, and effective generalization across datasets. These findings highlight the potential of graph-enhanced multi-stage modeling for early-stage yield prediction, offering a scalable solution for precision agriculture and intelligent crop management.

Keywords: maize yield prediction; multi-stage maize growth data; data imputation; hypergraph attention network; bipartite graph neural network.

Introduction

Maize, as one of the most important staple crops worldwide, plays a critical role in modern agricultural management, grain production planning, and global food security strategies (Iizumi *et al.*, 2013; Guan *et al.*, 2017). Accurately estimating maize yield prior to physiological maturity not only facilitates informed decisions regarding cultivation and harvesting schedules, but also enables timely responses to potential environmental stresses, thereby improving resource use efficiency (Subedi *et al.*, 2005). Throughout the maize growing season, the emergence, silking, and maturity stages are recognized as key determinants of final yield (*Farahani et al.*, 2011) (Fig. 1a). For instance, the emergence stage directly affects seedling vigor and canopy establishment and is highly sensitive to soil moisture conditions (Zheng *et al.*, 2011); the silking stage represents a critical window for pollination and reproductive development, during which the crop is particularly vulnerable to stresses such as drought (Saseendran *et al.*, 2014; Gheysari *et al.*, 2017). Consequently, yield prediction based on multiple phenological stages of maize is not only of theoretical interest but also holds substantial practical value for precision agriculture and food system resilience.

Despite substantial progress in maize yield modeling, current approaches face two major challenges. First, most studies focus on phenotypic traits collected at a single time point during the maturity stage, overlooking the dynamic evolution of traits across multiple growth stages that critically influence yield formation. This limits the temporal responsiveness and predictive foresight of such models (Peng *et al.*, 2020; Jägermeyr *et al.*, 2021). Second, large-scale agricultural datasets are often characterized by extensive missing or incomplete observations, a challenge that becomes more pronounced in multi-stage and cross-regional monitoring settings (Fig. 1b). Due to the uneven distribution of weather stations, interrupted observations, or human errors, data from early developmental stages (e.g., emergence) and from specific regions frequently suffer from substantial gaps, which significantly undermines the stability and generalizability of predictive models (Sakamoto *et al.* 2013; Shahhosseini *et al.* 2020; Yuan *et al.* 2022). In addition, strong environmental heterogeneity across geographic regions in large-scale agricultural systems (Fig. 1c) further exacerbates the dependency of predictive performance on spatial continuity and structured data availability.

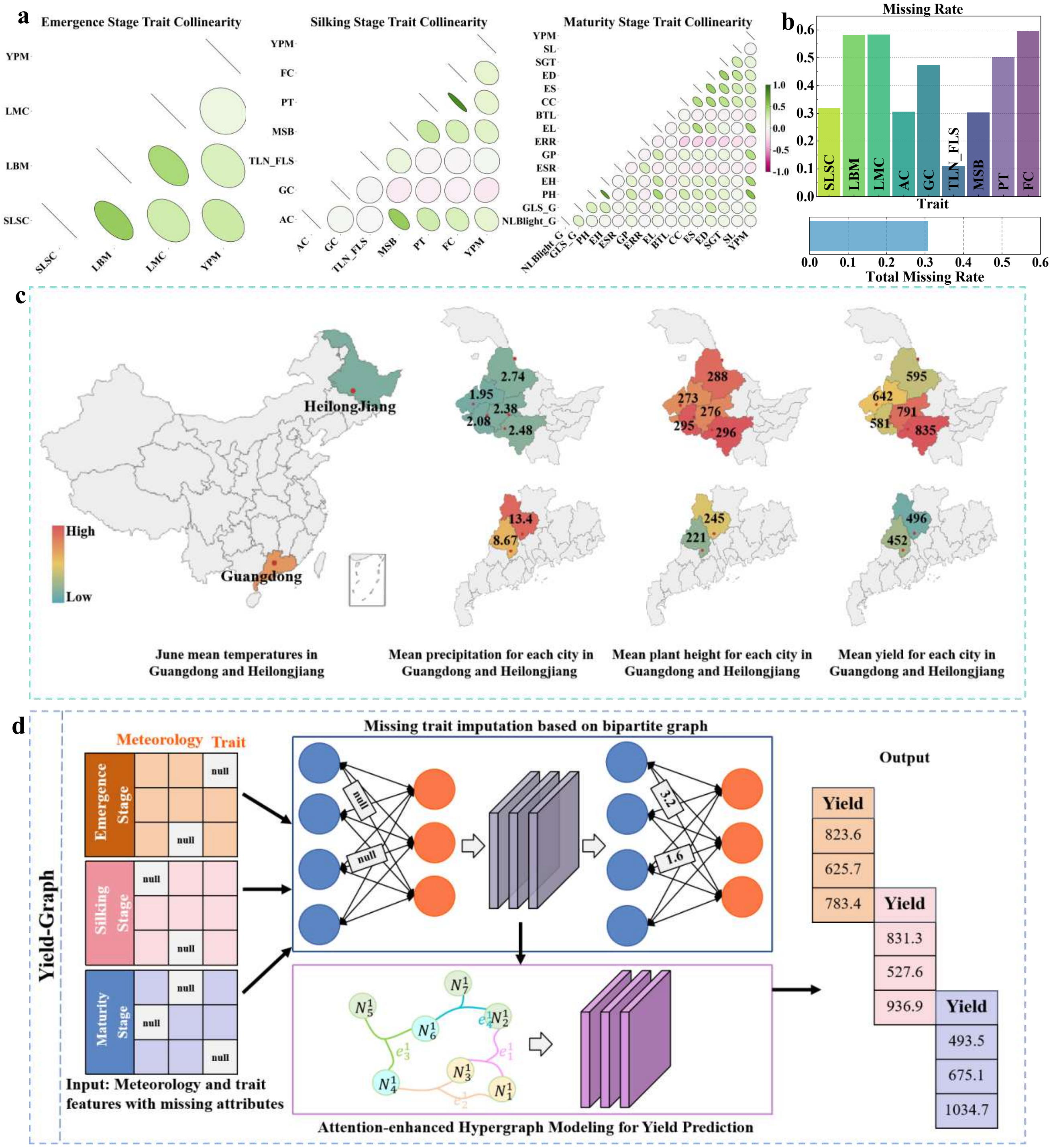

Fig. 1 Overview of trait correlation, missingness distribution, spatial feature patterns, and the proposed graph-based imputation and prediction framework. **a,** Pairwise collinearity analysis of agronomic traits during the emergence, silking and maturity stages. Green and pink cells represent positive and negative Pearson correlations, respectively. **b,** Missing rate of each trait variable and the overall missing ratio across all samples, highlighting the incomplete nature of multi-stage agronomic records. **c,** Spatial distribution of key agro-meteorological and phenotypic variables across cities in Guangdong and Heilongjiang provinces, including mean temperature (°C) in June, annual precipitation (mm), average plant height (cm), and grain yield per unit area (kg·mu^{-1}). Color gradients indicate the magnitude of each variable, with numerical annotations providing specific values. **d,** Schematic illustration of the Yield-Graph method. A bipartite graph neural network is used to impute missing trait values across three maize growth stages, followed by an attention-enhanced hypergraph neural network for final yield prediction.

The integration of remote sensing, multi-source environmental data, and data-driven modeling has been widely explored as a means to address these challenges (Matese et al. 2025; Karunathilake et al. 2025; Nayineni et al. 2025). For instance, Peng *et al.* (2020) developed a yield regression model based on solar-induced chlorophyll fluorescence (SIF) and vegetation indices, while

Rischbeck *et al.* (2016) integrated thermal imaging and spectral information to enhance yield prediction under varying climatic conditions. However, remote sensing approaches are often constrained by low temporal resolution, complex image preprocessing, and high sensitivity to environmental interference, limiting their capacity to monitor multi-stage trait variations at high frequency. In contrast, environmental variables—due to their fundamental and widespread availability—continue to serve as core features in yield modeling (Cane *et al.*, 1994; Carew *et al.*, 2009; Choquette *et al.*, 2023; Bowerman, 2024; Hogan *et al.*, 2024; Verrico *et al.*, 2025). Traditional statistical and machine learning approaches, such as random forests (Li *et al.*, 2021), regression models (Neto *et al.*, 2017; Gaona *et al.* 2023), and clustering techniques (Han *et al.*, 2016; Jung *et al.*, 2016), have been applied to integrate environmental features into predictive frameworks. Nevertheless, these methods often fall short in capturing cross-regional spatial dependencies and the temporal dynamics of trait evolution. In recent years, graph neural networks (GNNs) have emerged as a promising tool for representing spatial structures and sample relationships in agricultural data. Fan *et al.* (2021) proposed a recurrent neural network architecture based on graph structures for crop yield prediction, while Qiao *et al.* (2023) introduced the KSTAGE framework, which incorporates spatiotemporal dynamics. Despite these advances, most graph-based models still rely on single-stage inputs from maturity, failing to capture the influence of trait dynamics at earlier growth stages on yield outcomes. Moreover, their capacity for modeling higher-order relationships remains limited. This limitation becomes more pronounced when dealing with missing data and strong regional heterogeneity, where traditional graph structures struggle to capture the nonlinear, multi-entity interactions inherent in large-scale agricultural systems (Xu *et al.*, 2004). In this context, hypergraph models offer a natural advantage by explicitly modeling high-order interactions among multiple entities (Battiston *et al.*, 2020; Battiston *et al*., 2021; Bick *et al.*, 2023), making them particularly well-suited for representing the complex dependencies between environmental variables and crop traits.

To address the aforementioned challenges, we propose Yield-Graph, a yield prediction method based on multi-stage growth data of maize, which integrates bipartite graph modeling and a hypergraph attention mechanism to simultaneously capture trait completeness and spatial structural dependencies (Fig. 1d). Specifically, a sampl-trait bipartite graph is first constructed to impute missing phenotypic traits, leveraging the message-passing capabilities of the graph structure to enable spatial information sharing across planting regions (Hamilton *et al.*, 2017; You *et al.*, 2020; Cao *et al.*, 2021). Subsequently, a hypergraph is built based on trait similarity to capture high-order interactions among samples. An attention mechanism is introduced to enhance the modeling of inter-trait associations, thereby improving the accuracy and robustness of yield prediction (Bai *et al.*, 2021).

Experimental results demonstrate that the proposed Yield-Graph consistently outperforms existing models across multiple publicly available datasets, with particularly strong performance during the emergence, silking, and maturity stages—three critical phases in maize development. The method also achieves superior imputation accuracy for missing traits compared to other graph-based approaches. The proposed framework exhibits strong adaptability across growth stages and generalization across regions, offering a practical and scalable solution for crop yield prediction under complex environmental conditions.

Materials and methods

Collection Maize-PT data (maize planting and trait data)

The maize cultivation data set includes meteorological data from planting sites and phenotypic trait records throughout the growth cycle. Meteorological data, sourced from the National Meteorological Information Center of China, cover daily weather records across counties. Eleven key meteorological variables affecting maize growth, such as daily maximum temperature, mean pressure, and sunshine duration, were selected.

Maize yield prediction focuses on three growth stages: emergence, silking, and maturity. At each stage, 11 meteorological variables were analyzed, with their mean and variance computed, yielding 22 meteorological traits (Table 1). Phenotypic data, collected from 102 sites (2017-2022), include 10,000 records with some missing values. Growth stages are defined by specific developmental periods, with traits recorded accordingly (Table 2). Before bipartite graph representation, data preprocessing involved two steps: categorical variables (e.g., leaf blight resistance, cob color, ear shape) were numerically encoded, and numerical variables were standardized to remove scale effects and ensure comparability.

In addition to the above datasets, we also conducted yield prediction experiments using the publicly available rice phenotypic data set HeadingDate (Zhao et al., 2025). The selected traits and their corresponding abbreviations are summarized in Table 3.

Table 1: Meteorological traits and abbreviations

Feature Name	Abbreviation	Feature Name	Abbreviation
Maximum Temperature Mean	Tmax_Mean	Maximum Temperature Variance	Tmax_Var
Average Temperature Mean	Tavg_Mean	Average Temperature Variance	Tavg_Var
Minimum Temperature Mean	Tmin_Mean	Minimum Temperature Variance	Tmin_Var
Temperature Difference Mean	Tdiff_Mean	Temperature Difference Variance	Tdiff_Var
Ground Pressure Mean	P_Mean	Ground Pressure Variance	P_Var
Relative Humidity Mean	RH_Mean	Relative Humidity Variance	RH_Var
Precipitation Mean	Prec_Mean	Precipitation Variance	Prec_Var
Maximum Wind Speed Mean	WSmax_Mean	Maximum Wind Speed Variance	WSmax_Var
Average Wind Speed Mean	WSavg_Mean	Average Wind Speed Variance	WSavg_Var
Sunshine Duration Mean	SD_Mean	Sunshine Duration Variance	SD_Var
Wind Scale Mean	WScale_Mean	Wind Scale Variance	WScale_Var

Table 2: Maize trait and abbreviations

Stage	Trait Name	Abbreviation	Unit
Emergence Stage	Seedling Leaf Sheath Colour	SLSC	Score
	Leaf Blade Color	LBM	Score
	Leaf Margin Color	LMC	Score
Silking Stage	Anther Color	AC	Score
	Glume Color	GC	Score
	Total Leaf Number at Full Life Span	TLN_FLS	count
	Male Spike Branching	MSB	count
	Plant Type	PT	Score
	Filament Color	FC	Score
Maturity Stage	Northern Leaf Blight Grade	NLBlight_G	Grade
	Gray Leaf Spot Grade	GLS_G	Grade
	Plant Heigh	PH	cm
	Ear Heigh	EH	cm
	Empty Stem Rate	ESR	%
	Ear Length	EL	cm
	Bald Tip Length	BTL	%
	Cob Color	CC	Grade
	Ear Shape	ES	Grade
	Ear Diameter	ED	cm
	Stay-Green Trait	SGT	Grade
	Shank Length	SL	cm
	Growth Period	GP	d (days)
	Ear Rot Rate	ERR	%
	Yield per Mu	YPM	kg·mu^{-1}

Table 3: Rice Trait Characteristics and Abbreviations

Trait Name	Abbreviation	Trait Name	Abbreviation	Trait Name	Abbreviation
Heading Data	HD	Grain Length	GL	Grain Thickness	GT
Plant Height	PH	Grain Width	GW_h	Length to Width Ratio	LWR
Number of Panicles	NP	Grain Weight	GW_t	Yield	Y
Number of Effective Panicles	NEP	Spikelet Length	SL		

Preprocessing of Maize-PT data

For maize yield prediction across three growth stages, 11 meteorological features were selected for each stage, resulting in a total of 22 dimensions. The mean of each feature represents overall weather conditions, while variance reflects daily fluctuations. Two preprocessing steps were applied to both meteorological and maize planting data. First, categorical data (e.g., maize variety, seedling leaf sheath color) were converted into numerical labels, such as encoding axis color as 1 for white, 2 for red, 3 for pink, and 4 for purple. Second, to ensure comparability across features with varying value ranges, all numerical data were standardized, as shown in Eq. (1), minimizing any disproportionate influence on the prediction model.

$$x' = \frac{x-\mu_x}{\sigma_x} \tag{1}$$

where x denotes the data before processing, which μ_x is the mean value of this dimension of data, σ_x denotes the standard value of this dimension of data. The normalization procedure mitigates the potential distortion of the model due to large discrepancies between different feature values.

Yield-Graph model structure

This study presents Yield-Graph, a maize trait imputation and yield prediction framework that integrates bipartite graph neural structures with a hypergraph-based attention mechanism to enhance predictive accuracy under diverse environmental conditions. Traditional models are constrained by spatial dependencies, limiting their generalizability across different regions. To address this limitation, the proposed approach employs a progressive learning strategy to comprehensively capture the complex interactions between environmental factors and crop traits, thereby improving cross-regional predictive capability.

The detailed framework is illustrated in Fig. 2, the Yield-Graph framework consists of two parts. In the first part (Fig. 2a-d), data is transformed into a bipartite graph to model the relationships between environmental factors and crop traits. A bipartite graph update network is utilized for missing data imputation, effectively capturing high-order dependencies among variables. In the second part (Fig. 2e-g), the bipartite graph is further converted into a hypergraph, where an attention mechanism is incorporated into the hypergraph neural network to refine yield prediction accuracy.

First, treating meteorological feature names and maize trait feature names as one class of nodes (m dimensional, denoted as $F_j,\ j \in (1, m)$), and planting data identifiers as another class of nodes (n dimensional, denoted as $N_i,\ i \in (1, n)$). The table contains a total of $n \times m$ data values except for acres, so the bipartite graph contains $n \times m$ edges. The j^{th} feature of the i^{th} data can be represented by the edge e_{ij}, and the data value is the weight of e_{ij}. If the value is null, $e_{ij} = 0$. The data values serve as the edge weights. V_i^l denotes the ith target node in the lth feature update layer. j denotes the jth node among them. e_{ij}^l denotes the weight of the edge connecting node V_i^l to node V_j^l. During the graph construction phase, one-hot encoding was employed to distinguish between different types of nodes. This graph structure effectively captures the relationships between planting data and maize traits, providing a foundation for the subsequent imputation of missing data.

In the design of the graph neural network architecture, we propose a model comprising three graph update layers and one feature imputation layer. Each graph update layer consists of two sequential steps. The first step updates node features using edge embeddings. Notably, the node update follows a unified strategy that does not differentiate between sample nodes and feature nodes. The update is formulated as follows:

$$V_i^{l+1} = F_b\left(V_i^l \oplus [\sum_{j\in\mathcal{N}_i} F_a(V_i^l \oplus e_{ij}^l)\,]\right) \quad (2)$$

Here, $\mathcal{N}_i$ denotes the set of all neighboring nodes connected to node i, both $\mathrm{F_a}$ and $\mathrm{F_b}$ represent fully connected layers, and $\oplus$ denotes the concatenation operation. Subsequently, the model updates the edge weights based on the current layer's node features, thereby enhancing the discriminative power and representational capacity of the edges during feature propagation. The edge update is computed as follows:

$$e_i^{l+1} = (V_i^{l+1} \oplus V_j^{l+1} \oplus e_{ij}^l) \quad (3)$$

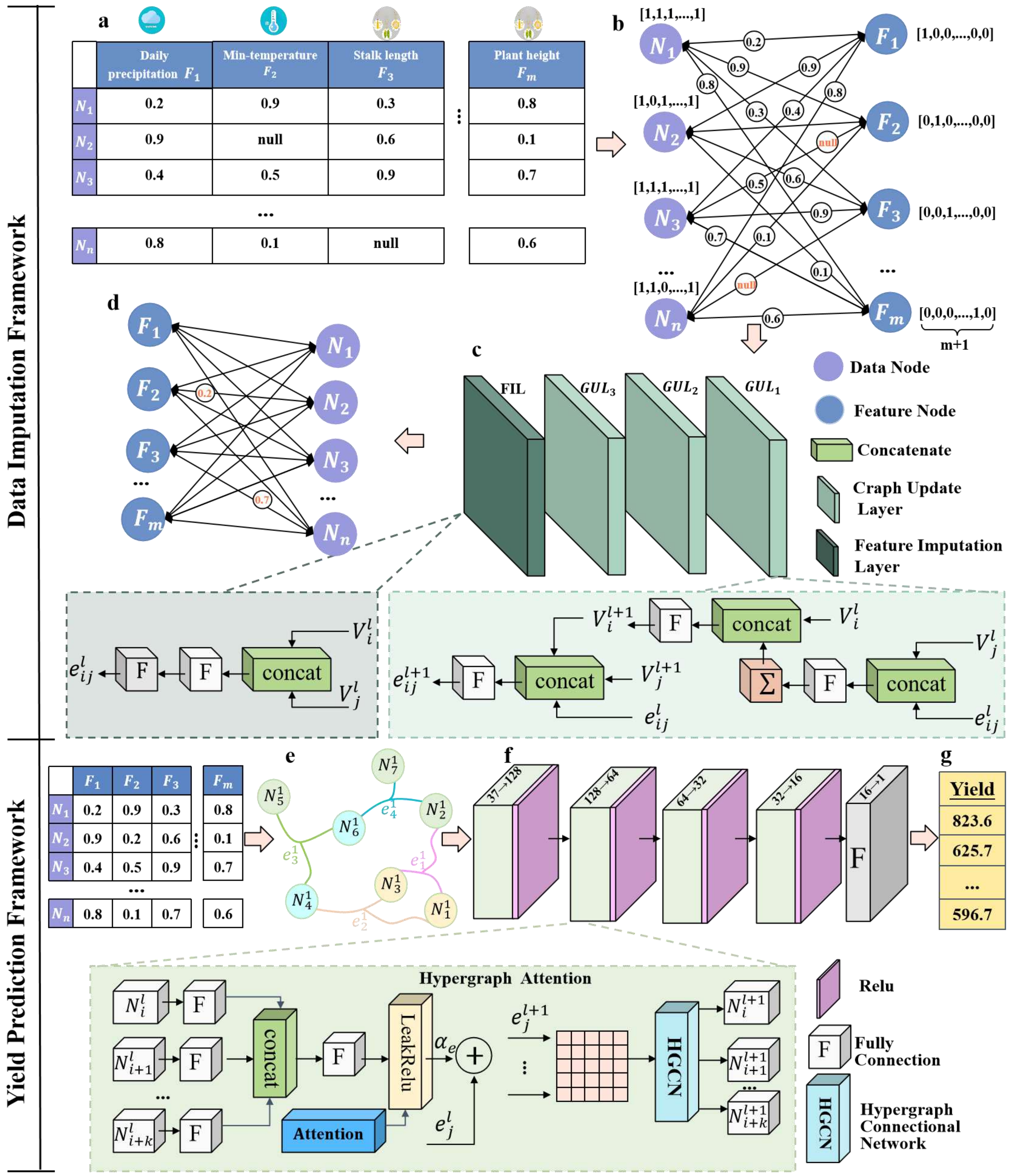

Fig. 2 Detailed Architecture of Maize Yield Prediction Framework Based on Bipartite Graph Network for Data Imputation and HyperGraph Attention Network. **(a-d)** Data Imputation Framework: This framework employs a bipartite graph structure for data imputation, where nodes represent planting data and features, and edges capture their associations. The network consists of three graph update layers (light green rectangles) and one feature imputation layer (dark green rectangle). **(e-g)** Maize Yield Prediction

Framework: This framework is constructed based on a hypergraph attention network, utilizing a hypergraph structure to model the complex interactions between maize growth environments and yield.

This design enables progressive integration of high-order structural information between nodes and edges across the three graph update layers, effectively capturing complex cross-variable dependencies among planting samples. Finally, in the feature imputation layer, the model takes the node features obtained from the third graph update layer as input and predicts the missing trait values through two fully connected layers (represented by the grey rectangles in Fig. 2c).

To capture the complex interactions among environmental conditions, maize trait features, and yield, we construct hyperedges using a K-nearest neighbors (KNN) strategy with $k = 3, 4$ and 5, where each hyperedge connects $k + 1$nodes. This design enhances the model's generalization capability under heterogeneous environmental conditions. The specific structure of the hypergraph attention layer is shown in the (light green dashed box in the lower half of Fig. 2f. It demonstrates the process of updating node features and hyperedge weights in the l^{th} hypergraph attention layer. Firstly, the hyperedge weights are updated based on node features connected to the hyperedge. The process is shown as:

$$e_j^{l+1} = e_j^l + LeakyRelu(L_{Full}(Con(L_1(N_i^l), \ldots, L_{Full}(N_{i+k}^l)))) \tag{4}$$

Where LeakyRelu is the activation function; e_j^{l+1} is the weight of the j^{th} hyperedge after the update of the l^{th} layer; N_i^l is the input feature of the i^{th} node of the l^{th} layer; k denotes a hyperedge connecting $k + 1$ nodes. After all hyperedges have finished updating, the hypergraph is represented by a new Laplace matrix. It is input to a HGCN to finish updating nodes. In Fig. 2, the input and output dimensions of each hypergraph attention layer for the node features are marked at the top of each light green rectangle. The last hypergraph attention layer outputs a node feature with 16 dimensions. It is input into a fully connection layer (gray rectangle) to get the yield prediction result.

Compared to traditional methods, this framework demonstrates significant advantages in handling missing data, modeling feature dependencies, and enhancing cross-environment prediction. It provides a robust scientific basis for precision agriculture, crop genetic improvement, and environmental regulation, with potential applications extending to wheat, rice, and soybean, facilitating data-driven intelligent agricultural management.

Evaluation metrics

The yield prediction model was evaluated using five metrics: the coefficient of determination (R^2, Eq. 5), mean square error (MSE, Eq. 6), root mean square error (RMSE, Eq. 7), mean absolute error (MAE, Eq. 8), and relative error (Eq. 9). R^2 quantifies the agreement between predicted and actual yields, with higher values indicating better fit. RMSE assesses data dispersion, while MSE and MAE measure overall prediction errors, reducing the impact of error cancellation. The definitions of these metrics are provided below:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{\sum_{i=1}^{n}(\overline{y_i} - y_i)^2} \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2 \tag{6}$$

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{8}$$

$$Relative\ error = \frac{\hat{y}_i - y_i}{y_i} \tag{9}$$

where n represents the number of data in a test set, y_i and $\widehat{y_i}$ represents the accurate corn yield and the predicted corn yield for the i^{th} test data sample, $\overline{y_i}$ represents the mean value of the reported corn yield.

Results

High-precision yield prediction driven by multi-stage maize growth in Yield-Graph

To evaluate the predictive performance of the proposed method, we conducted yield prediction experiments using a large-scale maize dataset comprising 10,000 samples. The dataset was divided into a training set (80%, i.e., 8,000 samples) and a test set (20%, i.e., 2,000 samples). To ensure a comprehensive comparison, we selected nine representative benchmark models, including five widely used machine learning algorithms—Random Forest, AdaBoost, Gradient Boosting, XGBoost (Chen *et al.*, 2016), and LightGBM (Ke *et al.*, 2017)—one deep learning model, TabNet (Arik *et al.*, 2021), and three graph-based neural network models: Graph Convolutional Network (GCN) (Kipf *et al.*, 2016), Graph Attention Network (GAT) (Veličković *et al.*, 2017), and Hypergraph Residual Network (HGRN) (Zhao *et al.*, 2018). Prior to model training, all approaches employed a unified bipartite graph-based imputation strategy to address missing trait values and maintain data consistency.

Yield prediction results for three critical growth stages of maize—emergence, silking, and maturity—are reported in Tables S1, S2, and S3, respectively. To further investigate the impact of the hyperedge connectivity parameter k on prediction performance, we evaluated the proposed Yield-Graph model under three configurations: k=3, k=4, and k=5, denoted as Our(k=3), Our(k=4), and Our(k=5), respectively. In all result tables, the best-performing values are highlighted in bold, and the second-best results are underlined.

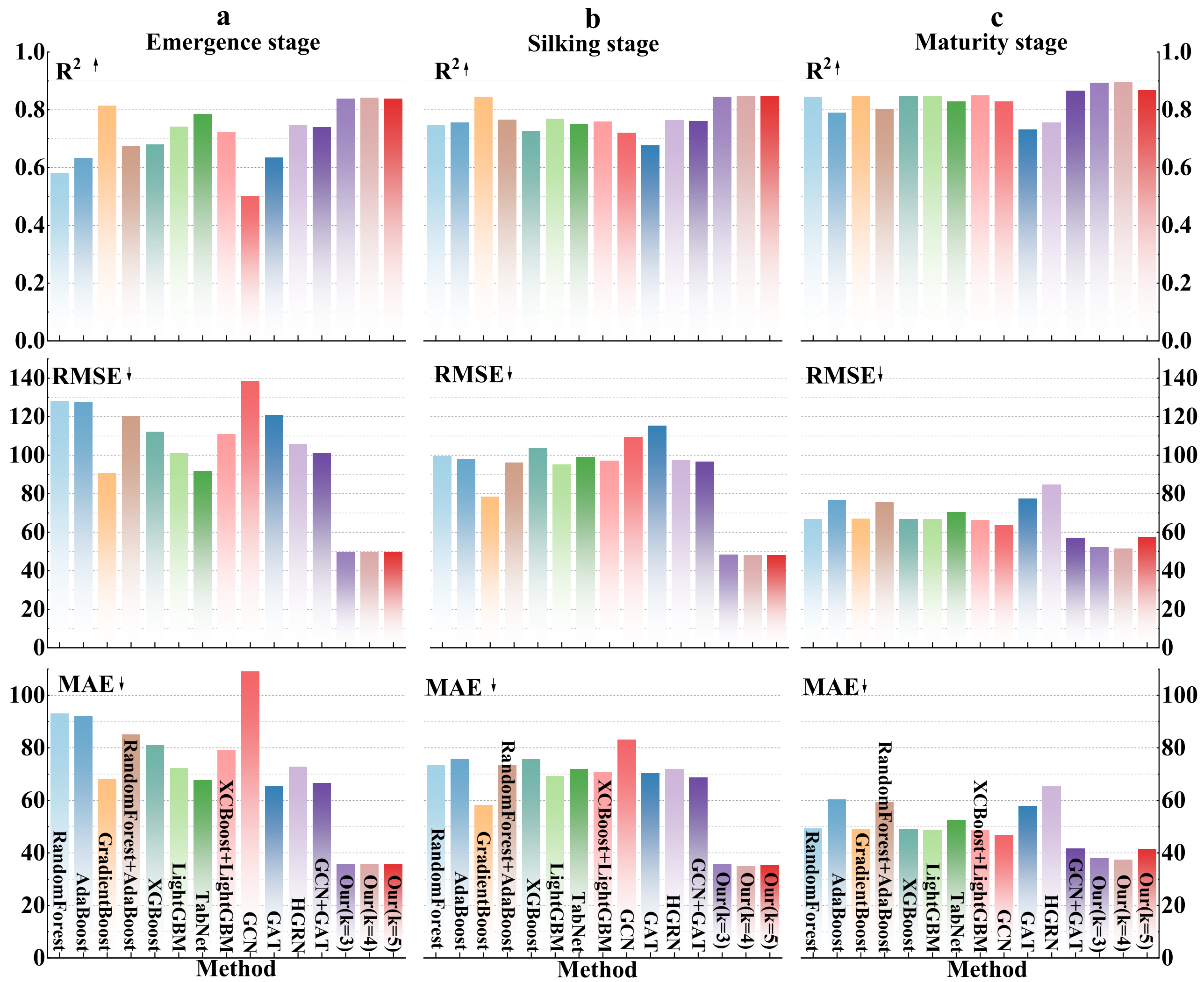

Fig. 3 Yield prediction performance of different models across three maize growth stages. **a**, Performance based on emergence stage data. **b**, Performance based on silking stage data. **c**, Performance based on maturity stage data.

At the emergence stage (Table S1, Fig. 3a), although ensemble models such as GCN and GAT demonstrated relatively good performance, the proposed Yield-Graph achieved more stable and accurate predictions under the configuration of k=4, with a coefficient of determination R^2=0.842. This result highlights the significant advantage of the hypergraph structure in capturing complex nonlinear relationships during early growth. During the silking stage (Table S2, Fig. 3b), machine learning models such as

Gradient Boosting and LightGBM yielded competitive results. However, the Yield-Graph model consistently outperformed all other methods across configurations k=3,4 and 5, with R^2 values reaching 0.848 under both k=4 and k=5. These findings further validate the robustness and effectiveness of the hypergraph neural network in modeling intricate phenotypic traits and gene–environment interactions.

At the maturity stage (Table S3, Fig. 3c), prediction performance improved across all models. While the GCN combined with GAT ensemble attained an R^2=0.866, the Yield-Graph with k=4 achieved a higher R^2=0.895, along with lower RMSE (51.51) and MAE (37.43), demonstrating superior predictive accuracy and stability. The improved performance at this stage may be attributed to clearer phenotypic patterns and higher data quality, which facilitate more accurate modeling. In contrast, performance during the emergence stage was limited by data uncertainty and missing values. As illustrated in Figs. 3a-c, the predictive results of all models were compared across the three critical maize growth stages: emergence, silking, and maturity. Each colored bar in the histogram represents a specific model. It is evident that Yield-Graph consistently exhibited superior performance across all stages, and particularly outperformed other approaches during the maturity stage. These results confirm the effectiveness of Yield-Graph in modeling yield variation throughout multiple growth stages, underscoring its potential as a reliable tool for high-precision crop yield forecasting.

Accuracy of yield predictions across three developmental stages

Accurate maize yield forecasting plays a critical role in agricultural production management and resource allocation. However, most existing predictive models rely heavily on highly complete datasets and are often restricted to specific spatial regions, thereby limiting their generalizability. To overcome these challenges, we propose Yield-Graph, a novel prediction framework that integrates a bipartite graph neural network for imputing missing phenotypic attributes with a hypergraph attention network for modeling spatiotemporal dependencies. Notably, this approach maintains predictive accuracy even in the presence of incomplete trait information.

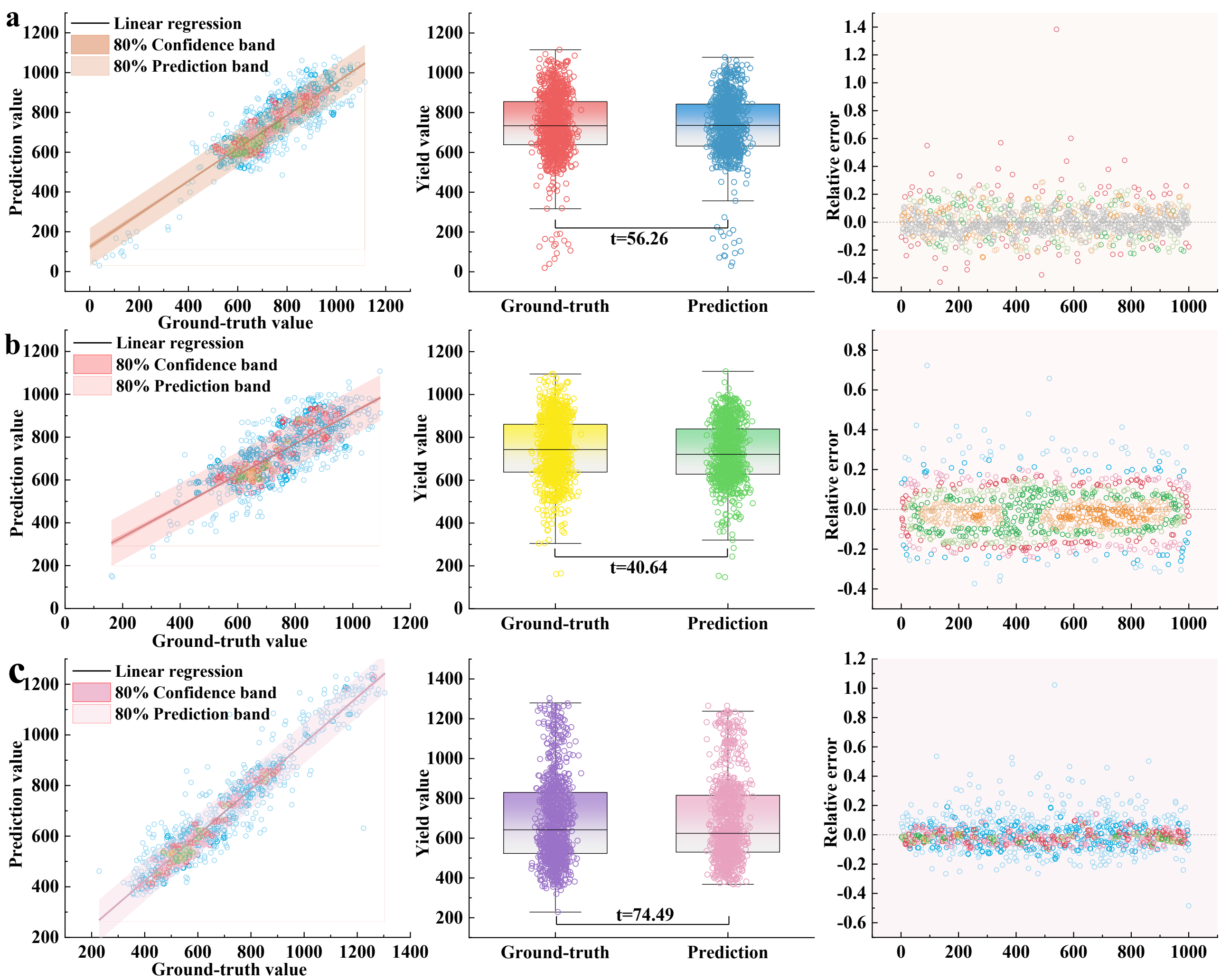

Fig. 4 Comparison between predicted and actual maize yield across growth stages and spatial regions. **a**, From left to right: regression plot of predicted vs. actual yield, boxplot comparison, and relative error distribution for the emergence stage. **b**, Corresponding regression, boxplot, and error distribution for the silking stage. **c**, Corresponding regression, boxplot, and error distribution for the maturity stage.

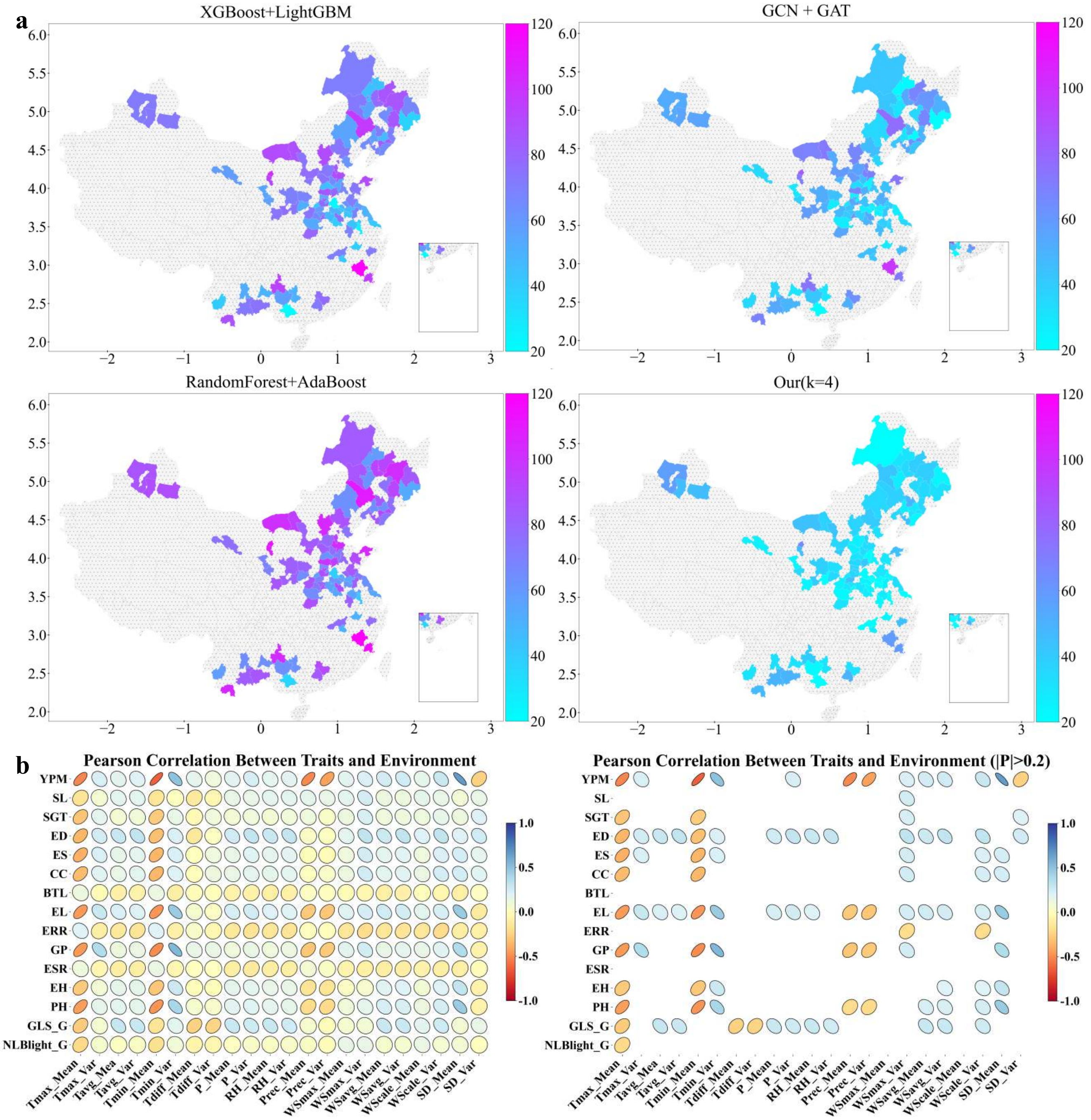

Fig. 5 Spatial prediction error and trait–environment correlation analysis for maize yield at maturity stage. **a**, Spatial distribution of yield prediction errors comparing Yield-Graph (our(k = 4)) with various ensemble models. Red indicates higher errors, while blue indicates lower errors; Yield-Graph demonstrates lower errors in major maize-producing regions, reflecting superior spatial generalization. **b**, Pearson collinearity feature selection (PCS) analysis between traits and environmental variables for maize yield prediction at maturity stage. The shape and color of each ellipse represent the strength and direction of correlation: deeper color and more flattened ellipses indicate stronger correlations, with blue for positive and red for negative associations. Filtered correlation map showing significant trait–environment relationships ($|r| > 0.2$ and $p < 0.05$), retaining the most informative interactions. Collinearity analysis among environmental variables, with green indicating positive and pink indicating negative correlations.

To systematically evaluate model performance, yield predictions were validated at three key developmental stages of maize growth—emergence, silking, and maturity (Fig. 4a-c). Approximately 80% of the predicted values showed strong agreement with observed yields in the scatter plots. Boxplots and error histograms further indicated low variability and stable distributions of the predictions. Significance tests between predicted and observed yields confirmed significant correlations at all three stages. Notably, while yield predictions based on traits from earlier stages provided only coarse estimates, predictions at the maturity stage achieved the highest accuracy, highlighting that phenotypic traits and environmental conditions during this period exert the greatest influence on final yield. Spatial error mapping (Fig. 5a) compared Yield-Graph (k=4) against ensemble methods (RandomForest+AdaBoost, XGBoost+LightGBM, GCN+GAT). While baselines showed spatially dispersed errors, Yield-Graph achieved lower, geographically consistent errors in core maize-producing regions, confirming its superior spatiotemporal generalization.

To elucidate trait–environment interactions, Pearson correlation-based screening (PCS) was applied to 15 phenotypic traits and 22 environmental variables (Fig. 5b). Yield per mu (YPM) was positively correlated with mean solar duration (SD_Mean) and negatively with minimum temperature (Tmin_Mean), suggesting that enhanced light and reduced nocturnal cooling promote yield. Disease traits (e.g. NLBlight_G, GLS_G) showed negative correlations with maximum temperature, implying lower disease pressure in warmer climates. These findings reinforce Yield-Graph's capacity to model complex agro-ecological patterns and guide both breeding and management decisions.

Evaluation of cross-species generalization: rice yield prediction

To comprehensively assess the cross-species generalization capacity of the proposed Yield-Graph framework, we conducted yield prediction experiments on the publicly available rice phenotypic dataset HeadingDate (Zhao *et al.*, 2025). The selected traits and their corresponding abbreviations are listed in Table 3. The results indicate that Yield-Graph maintains strong applicability and robustness when applied to rice, a crop with substantially different phenotypic profiles compared to maize.

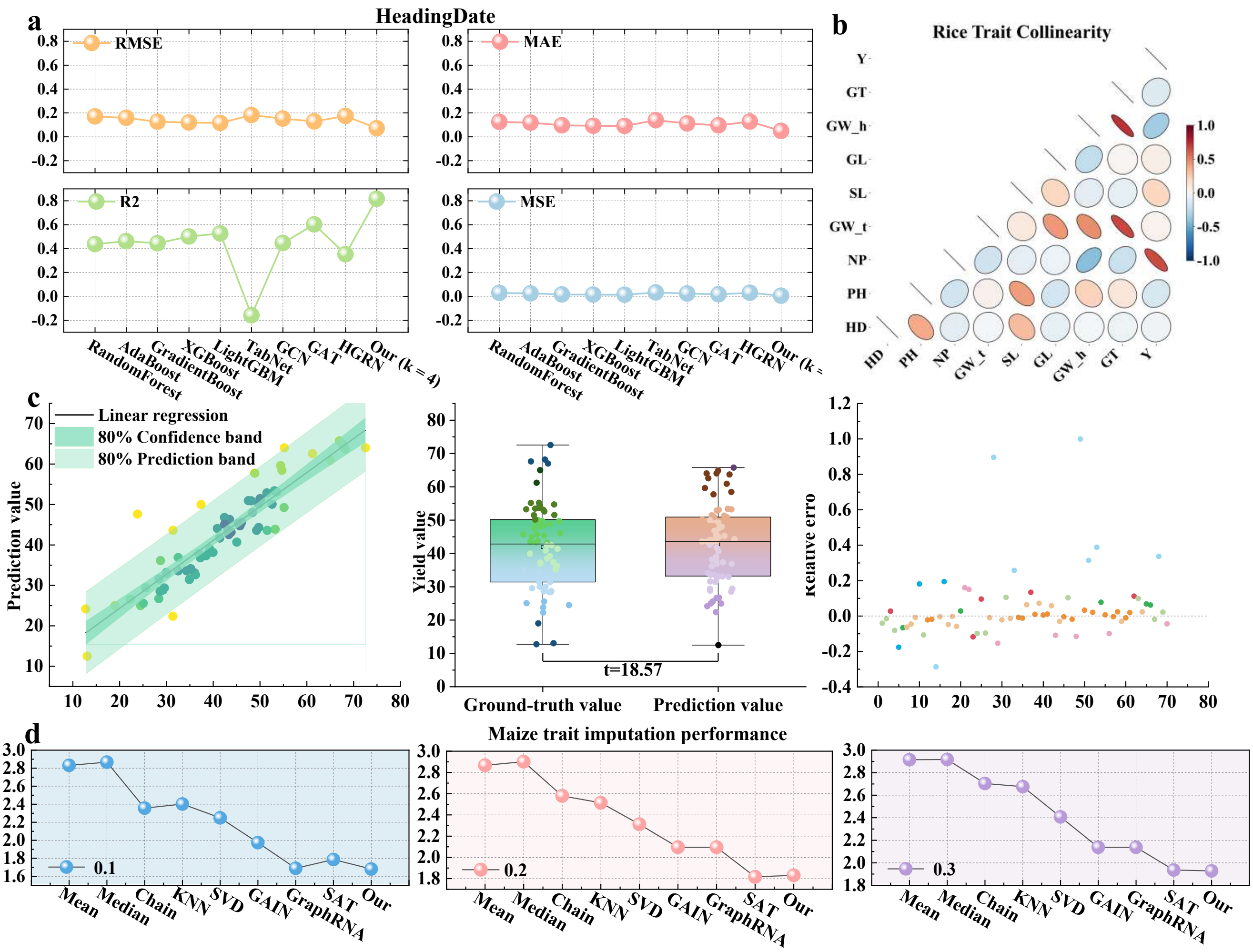

Fig. 6 Visualization and analysis of rice yield prediction and imputation performance. **a**, Performance comparison of different models for rice yield prediction, highlighting the generalization capability of Yield-Graph on cross-species data. **b**, Heatmap of

pairwise Pearson correlations among key phenotypic traits. The shape and color of each ellipse indicate the strength and direction of correlation: deeper colors and more elongated ellipses represent stronger correlations; red denotes positive and blue denotes negative correlation. **c**, Regression plot of predicted vs. actual yield values, boxplot comparison, and relative error distribution. **d**, Evaluation of the imputation performance of different methods under varying levels of missing trait data in maize.

As illustrated in Fig. 6a (Table S4), Yield-Graph consistently outperforms benchmark methods across multiple evaluation metrics, achieving higher R^2 values and lower MAE, MSE, and RMSE scores. These outcomes highlight its strong modeling capability and generalization performance in handling heterogeneous, multi-source phenotypic data. Furthermore, to examine the influence of individual phenotypic traits on rice yield, we performed Pearson correlation analysis (Fig. 6b). The analysis reveals that, aside from the number of panicles (NP) and effective panicles (NEP), the leaf width ratio (LWR) exhibits a significant positive correlation with yield, suggesting its potentially critical role in rice yield formation.

Fig. 6c provides a multi-faceted validation of the model's predictive accuracy and robustness. The left panel shows a scatter plot of predicted versus observed values, with data points closely aligned along the regression line, indicating good model fit. he middle panel shows the boxplot distributions of predicted and observed values, which exhibit a high degree of consistency and significant correlation, demonstrating that the model captures overall trends while remaining adaptive to extreme values. The right panel depicts the distribution of relative prediction errors, which are largely concentrated around zero, confirming the model's high precision and controlled error variance.

Collectively, these findings affirm that the proposed Yield-Graph model not only achieves superior performance in maize yield prediction but also demonstrates strong cross-crop generalization when applied to rice, highlighting its broad applicability and potential utility in diverse agricultural contexts (Figs. 6a-c).

Evaluation of imputation effectiveness for missing phenotypic traits

To address widespread missing phenotypic data in large-scale maize datasets, we evaluated our graph-based imputation method using 10,000 samples with simulated missingness levels of 10%, 20%, and 30% (mean missing rate ≈30%; Fig. 1b). Mean absolute error (MAE) was used to assess imputation accuracy (Fig. 6d). We compared our bipartite graph method with eight established techniques, including statistical (mean, median), regression-based (chained regression, KNN), matrix factorization (SVD), generative (GAIN), and graph neural network models (GraphRNA, SAT).

Statistical methods performed poorly across all conditions. While regression and matrix-based models maintained moderate accuracy under low missingness, their performance declined sharply as missingness increased. GAIN outperformed traditional methods by capturing nonlinear patterns. Among GNNs, SAT and GraphRNA exhibited robustness, but our bipartite graph model consistently achieved the lowest MAE at both low and high missingness levels.

By explicitly modeling inter-sample relationships, our method effectively recovered missing trait values and improved data integrity. In downstream yield prediction tasks, Yield-Graph demonstrated strong generalization and outperformed baselines across multiple metrics. These findings highlight the utility of graph-driven frameworks in enhancing data completeness and predictive accuracy under real-world conditions with incomplete phenotypic observations.

To address the prevalent issue of missing phenotypic trait data in large-scale maize cultivation datasets, we conducted a systematic evaluation using a dataset comprising 10,000 samples. As shown in Fig. 1b, the missing rates for individual traits ranged from 10% to 60%, with an average missing rate close to 30%. To quantitatively assess the performance of the proposed graph-based imputation method, we simulated three levels of missingness—10%, 20%, and 30%—and performed corresponding imputation procedures. The mean absolute error (MAE) was employed as the evaluation metric (Fig. 5d; blue, red, and purple lines indicate the three missingness levels, respectively).

Discussion

In this study, we proposed Yield-Graph, a unified graph-driven framework that simultaneously addresses missing phenotypic data and multi-stage yield prediction in maize. By integrating bipartite graph-based imputation with a residual hypergraph neural network, the model achieved superior imputation accuracy under high missingness and consistently outperformed ensemble learners, deep models, and existing GNNs in yield forecasting. Yield-Graph also demonstrated strong adaptability across growth stages, geographic regions, and even cross-crop validation on rice, highlighting its robustness and transferability. These results underscore the advantage of leveraging graph structures to capture complex, non-linear dependencies in agricultural data. Nonetheless, future work should incorporate genomic and environmental sensor features, as well as uncertainty-aware strategies for non-random missingness, to further enhance model generalizability in complex agro-ecosystems.

Conclusion

This study proposes the Yield-Graph model, which integrates bipartite graph-based imputation with a hypergraph attention mechanism to effectively capture the dynamic evolution of multi-stage phenotypic traits and high-order relationships among samples. The model demonstrates significant advantages in both trait completion and yield prediction tasks. It achieves high prediction accuracy across multiple maize growth stages and geographic regions, while also exhibiting strong cross-crop generalizability as validated on a rice dataset. These findings highlight the potential of Yield-Graph as a robust solution to challenges posed by incomplete phenotypic data and regional heterogeneity, offering promising applications for intelligent agricultural management and crop production decision-making.

Experimental equipment

In this paper, all experiments were conducted on Ubuntu 20.04 LTS with CUDA 12.2. The system utilizes an NVIDIA RTX A6000 48GB graphics card for training and testing . The initial learning rate was set to 0.015, with a maximum of 30,000 training epochs. An early stopping criterion was applied when the training loss dropped below 0.001 to prevent overfitting. The model was optimized using the Adam optimizer, with a MultiStepLR learning rate scheduler that decayed the learning rate at epochs 500, 2000, 4000, 6000, 10000, 15000, and 20000, using a decay factor of 0.9. To improve training efficiency and reduce memory consumption, automatic mixed precision (AMP) training was employed via torch.cuda.amp.

Statements & Declarations

Funding

This work is supported by the National Key Research and Development Program of China (No. 2024YFD1201500).
The authors have no relevant financial or non-financial interests to disclose.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.
On behalf of all authors, the corresponding author states that there is no conflict of interest.

Author contributions

All authors contributed to the conception and design of the study. Material preparation and data collection were performed by Kaiyi Wang and Xiangyu Zhao. Data analysis was carried out by Jiahui Wang, Yuqing Zhang, Yong Zhang, and Bo Li. The first draft of the manuscript was written by Jiahui Wang and Yuqing Zhang, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability

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

Code availability

The code developed to generate the results and analysis in this article is available at https://github.com/wjhhh2928/Yield-Graph

Supplementary information

Table. S1 Predicting maize yields from emergence stage.

Table. S2 Predicting maize yields from silking stage.

Table. S3 Predicting maize yields from maturity stage.

Table S4: Predicting rice yields

ORCID

Jiahui Wang ID: https://orcid.org/0009-0004-8398-9869

Yong Zhang ID: https://orcid.org/0000-0001-6650-6790

Kaiyi Wang ID: https://orcid.org/0009-0000-5365-7178

References

Anselin L, Bongiovanni R, Lowenberg-DeBoer J (2004) A spatial econometric approach to the economics of site-specific nitrogen management in corn production. Am. J. Agric. Econ. 86: 675-687. https://doi.org/10.1111/J.0002-9092.2004.00610.X

Arik SÖ, Pfister T (2021) TabNet: Attentive interpretable tabular learning. *In Proc. AAAI Conf. Artif. Intell.* 35: 6679-6687. https://doi.org/10.1609/aaai.v35i8.16826

Battiston F, Cencetti G, Lacopini L, Latora V, Lucas M, Patania A, Young JG, Petri G (2020) Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep* 874: 1-92. https://doi.org/10.1016/j.physrep.2020.05.004

Bai S, Zhang F, Torr P-S (2021) Hypergraph convolution and hypergraph attention. *Pattern Recognit* 110: 107637. https://doi.org/10.1016/j.patcog.2020.107637

Battiston F, Amico E, Barrat A, Bianconi G, Arruda G-F, Franceschiello B, Lacopini L, Kéfi S, Latora V, Murray M-M *et al*. (2021) The physics of higher-order interactions in complex systems. *Nat. Phys*. 17: 1093-1098. https://doi.org/10.1038/s41567-021-01371-4

Bick C, Gross E, Harrington H-A, Schaub M-T (2023) What are higher-order networks? *SIAM Rev.* 65: 686-731. https://doi.org/10.1137/21M1414024

Bowerman AF (2024) A demonstration of the enviromics approach to integrating environmental 'big data'problems. *New Phytologist* 245: 9-10. https://doi.org/10.1111/nph.20079

Cane M, Eshel G, Buckland R (1994) Forecasting Zimbabwean maize yield using eastern equatorial Pacific sea surface temperature. *Nature* 370: 204-205. https://doi.org/10.1038/370204a0

Carew R, Smith EG, Grant C (2009) Factors influencing wheat yield and variability: Evidence from Manitoba, Canada. *J. Agric. Appl. Econ.* 41: 625-639. https://doi.org/10.1017/S1074070800003114

Cao JX, Lin XX, Guo S, Liu LC, Liu TW, Wang B (2021) Bipartite graph embedding via mutual information maximization. *In Proc. 14th ACM Int. Conf. Web Search Data Min. (WSDM'21),* 635-643. https://doi.org/10.1145/3437963.344178

Chen X, Chen S, Yao J, Zheng H, Zhang Y, Tsang IW (2020) Learning on attribute-missing graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* 44: 740-757. https://doi.org/10.1109/TPAMI.2020.3032189

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. *In Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 785-794. https://doi.org/10.1145/2939672.2939785

Chang Y, Latham J, Licht M, Wang L (2023) A data-driven crop model for maize yield prediction. *Commun. Biol.* 6: 39. https://doi.org/10.1038/s42003-023-04833-y

Choquette NE, Holland JB, Weldekidan T, Drouault J, Leon N, Flint-Garcia S, Lauter N, Murray SC, Xu WW, Wisser RJ (2023) Environment-specific selection alters flowering-time plasticity and results in pervasive pleiotropic responses in maize. *New Phytologist* 238: 737-749. https://doi.org/10.1111/nph.18769

Farahani HA, Moaveni P, Maroufi K (2011) Effect of hydropriming on germination percentage in sunflower (Helianthus annuus L.) cultivars. *Adv. Environ. Biol.* 5: 2253-2258. https://link.gale.com/apps/doc/A267421993/AONE?u=googlescholar&sid=bookmark-AONE&xid=e4b97d1f

Fan J, Bai J, Li Z, Ortiz-Bobea A, Gomes CP (2021) A GNN-RNN approach for harnessing geospatial and temporal information: application to crop yield prediction. *In Adv. Neural Inf. Process. Syst.* 29. https://doi.org/10.1609/aaai.v36i11.21444

Gaona J, Verdugo PB, Fernández JM, Zamora AG, Martín LA, Jiménez CH (2023) Predictive value of soil moisture and concurrent variables in the multivariate modelling of cereal yields in water-limited environments. *Agric. Water Manag.* 289: 1-13. https://doi.org/10.1016/j.agwat.2023.108280

Gheysari M, Sadeghi SH, Loescher HW, Amiri S, Zareian MJ, Majidi MM, Asgarinia P, Payero JO (2017) Comparison of deficit irrigation management strategies on root, plant growth and biomass productivity of silage maize. *Agric. Water Manag.* 182: 126-138. https://doi.org/10.1016/j.agwat.2016.12.014

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial networks. *Commun. ACM* 63: 139-144. https://doi.org/10.1145/3422622

Guan K, Wu J, Kimball JS, Anderson MC, Frolking S, Li B, Hain CR, Lobell, DB (2017) The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. *Remote Sens. Environ*. 199: 333-349. https://doi.org/10.1016/j.rse.2017.06.043

Han L, Haleem MS, Taylor M (2016) Automatic detection and severity assessment of crop diseases using image pattern recognition. *In Emerg. Trends Adv. Technol. Comput. Intell., Springer*, 283-300. https://doi.org/10.1007/978-3-319-33353-3_15

Haghighattalab A, Crain J, Mondal S, Rutkoski J, Singh RP, Poland J (2017) Application of geographically weighted regression to improve grain yield prediction from unmanned aerial system imagery. *Crop Sci.* 57: 2478-2489. https://doi.org/10.2135/cropsci2016.12.1016

Hamilton WL, Ying Z, Leskovec J (2017) Inductive representation learning on large graphs. *In Adv. Neural Inf. Process. Syst.* 30: 1025-1035. https://dl.acm.org/doi/abs/10.5555/3294771.3294869

Hogan D, Schlenker W (2024) Non-linear relationships between daily temperature extremes and US agricultural yields uncovered by global gridded meteorological datasets. *Nat. Commun.* 15: 4638. https://doi.org/10.1038/s41467-024-48388-w

Huang X, Song Q, Li Y, Hu X (2019) Graph recurrent networks with attributed random walks. *In Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 732-740. https://doi.org/10.1145/3292500.3330941

Iizumi T, Sakuma H, Yokozawa M, Luo JJ, Challinor AJ, Brown ME, Sakurai G, Yamagata T (2013) Prediction of seasonal climate-induced variations in global food production. *Nat. Clim. Chang*. 3: 904-908. https://doi.org/10.1038/nclimate1945

Jägermeyr J, Müller C, Ruane AC, Elliott J, Balkovic J, Castillo O, Babacar F, Foster I, Folberth C, Franke JA *et al*. (2021） Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat. Food* 2: 873-885. https://doi.org/10.1038/s43016-021-00400-y

Jung I, Ahn H, Shin SJ, Kim J, Kwon HB, Jung W, Kim S (2016) Clustering and evolutionary analysis of small RNAs identify regulatory siRNA clusters induced under drought stress in rice. *BMC Syst. Biol*. 10: 423-432. https://doi.org/10.1186/s12918-016-0355-3

Karunathilake E.M.B.M., Thai TT, Mansoor S. Le AT, Baloch FS, Chung YS, Kim DW (2025) The use of RGB vegetation indices to predict the buckwheat yield at the flowering stage. J. Crop Sci. Biotechnol. 28, 379–389 . https://doi.org/10.1007/s12892-025-00283-6

Ke GL, Meng Q, Finley T, Wang TF, Chen W, Ma WD, Ye QW, Liu TY (2017) LightGBM: A highly efficient gradient boosting decision tree. *In Adv. Neural Inf. Process. Syst.* 30. https://api.semanticscholar.org/CorpusID:3815895

Khaki S, Wang L, Archontoulis SV (2020) A CNN-RNN framework for crop yield prediction. *Front. Plant Sci.* 10: 1-13. https://doi.org/10.3389/fpls.2019.01750

Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. *arXiv* preprint arXiv:1609.02907. https://api.semanticscholar.org/CorpusID:3144218

Klompenburg TV, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* 177: 105709. https://doi.org/10.1016/j.compag.2020.105709

Koren Y (2008) Factorization meets the neighborhood: a multifaceted collaborative filtering model. *In Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 426-434. https://doi.org/10.1145/1401890.140194

Li LC, Wang B, Feng PY, Wang HH, He QS, Wang YK, Liu DL, Li L, He JQ, Feng H *et al*. (2021) Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China. *Agric. For. Meteorol.* 308-309: 1-12. https://doi.org/10.1016/j.agrformet.2021.108558

Matese A, Hamie N, Baronti S *et al.* (2025) Integrating UAV hyperspectral imaging with machine learning techniques to predict tomato ecophysiological parameters and yield. *Precision Agric* 26, 72. https://doi.org/10.1007/s11119-025-10269-7

Nayak HS, McDonald AJ, Kumar V, Craufurd P, Dubey SK, Nayak AK, Parihar CM, Peramaiyan P, Poonia S, Tesfaye K *et al*. 2024. Context-dependent agricultural intensification pathways to increase rice production in India. *Nat. Commun.* 15: 8403. https://doi.org/10.1038/s41467-024-52448-6

Nayineni KK, Ragam P, Ahalya I, Jubilson EA, Viharika TVS, Sankati J (2025) Estimation of Crop Yield Using Intelligent Machine Learning Approaches. *In: Das, S.K., Behera, S.K. (eds) Recent Trends in Intelligent Systems and Next Generation Wireless Communication. IIWCS 2024. Lecture Notes in Networks and Systems,* 1329: 49-58. https://doi.org/10.1007/978-981-96-4741-5_5

Neto RJ, Souza ZM, Oliveira SM, Kölln OT, Ferreira DA, Carvalho JN, Braunbeck OB, Franco HJ (2017) Use of the decision tree technique to estimate sugarcane productivity under edaphoclimatic conditions. *Sugar Tech* 19: 662-668. https://doi.org/10.1007/s12355-017-0509-7

Peng B, Guan K, Zhou W, Jiang C, Frankenberg C, Sun Y, He LY, Köhler, P (2020) Assessing the benefit of satellite-based solar-induced chlorophyll fluorescence in crop yield prediction. *Int. J. Appl. Earth Obs. Geoinf.* 90: 102126. https://doi.org/10.1016/j.jag.2020.102126

Qiao MJ, He XH, Cheng XJ, Li PL, Zhao QB, Zhao CL, Tian ZH (2023) KSTAGE: A knowledge-guided spatial-temporal attention graph learning network for crop yield prediction. *Inf. Sci.* 619: 19-37. https://doi.org/10.1016/j.ins.2022.10.112

Rischbeck P, Elsayed S, Mistele B, Barmeier G, Heil K, Schmidhalter U (2016) Data fusion of spectral, thermal and canopy height parameters for improved yield prediction of drought stressed spring barley. *Eur. J. Agron.* 78: 44-59. http://dx.doi.org/10.1016/j.eja.2016.04.013

Sakamoto T, Gitelson A-A, Arkebauer TJ (2013) MODIS-based corn grain yield estimation model incorporating crop phenology information. *Remote Sens. Environ.* 131: 215-231. https://doi.org/10.1016/j.rse.2012.12.017

Saseendran SA, Ahuja LR, Ma L, Nielsen DC, Trout TJ, Andales AA, Chávez JL, Ham J (2014) Enhancing the water stress factors for simulation of corn in RZWQM2. *Agron. J.* 106: 81-94. https://doi.org/10.2134/agronj2013.0300

Shahhosseini M, Hu G, Archontoulis SV (2020) Forecasting corn yield with machine learning ensembles. *Front. Plant Sci.* 11: 1120. https://doi.org/10.3389/fpls.2020.01120

Subedi KD, Ma BL (2005) Seed priming does not improve corn yield in a humid temperate environment. *Agron. J.* 97: 211-218. https://doi.org/10.2134/agronj2005.0211a

Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2017) Graph attention networks. *arXiv* https://doi.org/10.48550/arXiv.1710.10903

Verrico B, Preston JC (2025) Historic rewiring of grass flowering timepathways and implications for cropimprovement under climate change. *New Phytologist* 245: 1864-1878. https://doi.org/10.1111/nph.20375

Xu N, York K, Miller P, Cheikh N (2004) Co-regulation of ear growth and internode elongation in corn. *Plant Growth Regul.* 44: 231-241. https://doi.org/10.1007/s10725-004-5935-3

Ying R, You JX, Morris C, Ren X, Hamilton WL, Leskovec J (2018) Hierarchical graph representation learning with differentiable pooling. *In Adv. Neural Inf. Process*, 31: 4805-4815. https://dl.acm.org/doi/10.5555/3327345.3327389

You J, Ma X, Ding Y, Kochenderfer MJ, Leskovec J (2020) Handling missing data with graph representation learning. *In Adv. Neural Inf. Process. Syst.* 33: 19075-19087. https://dl.acm.org/doi/abs/10.5555/3495724.3497325

Yuan Y, Lin F, Maucieri C, Zhang Y (2022) Efficient irrigation methods and optimal nitrogen dose to enhance wheat yield, inputs efficiency and economic benefits in the North China Plain. *Agronomy* 12: 273. https://doi.org/10.3390/agronomy12020273

Zheng ZP, Liu XH, Huang YB, Wu X, He C, Li Z (2011) A quantitative trait locus for the number of days from sowing to seedling emergence in maize. *Afr. J. Biotechnol.* 10: 5091-5095. https://academicjournals.org/journal/AJB/article-full-text-pdf/58C2B6130525

Zhao LY, Tang P, Lou JJ, Liu JX, Peng X, Shen MY, Wang CR, Zhao JL, Zhou DG, Fan ZL *et al.* (2025) Genomic prediction with NetGP based on gene network and multi-omics data in plants. *Plant Biotechnology Journal* 23: 1190-1201. https://doi.org/10.1111/pbi.14577 Dataset: https://figshare.com/articles/dataset/NetGP/26495956/4

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformation.pdf