Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN ADAPTIVE MULTI-SOURCE SPATIO-TEMPORAL GRAPH TRANSFORMER FRAMEWORK WITH CONTEXT-AWARE FEATURE FUSION FOR AGRICULTURAL YIELD FORECASTING AND FOOD SECURITY

JAGADISH A N[1], DR. AKKU MADHUSUDHAN[2], MRS. KARUNA VALISAMMAGARI[3], DR. POTNURI SURIBABU[4], RAMYA R[5], VAKAIMALAR ELAMARAN[6]

[1] Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, P O Box 325, Shamnur Road, Davanagere, Karnataka 577002, India
[2] Assistant Professor, Department of Electronics and Communication Engineering, Anurag University, Venkatapur, Ghatkesar, Hyderabad, Telangana 500088, India
[3] Assistant Professor, Department of Computer Science and Engineering (AI & ML), St. Peter's Engineering College, St. Peters College Road, Opposite TS Forest Academy, Dullapally, Maisammaguda, Medchal, Hyderabad, Telangana 500100, India
[4] Professor, Department of Computer Science and Engineering, DNR College of Engineering, Bhimavaram, West Godavari District, Andhra Pradesh 534201, India
[5] Department of Artificial Intelligence and Data Science, Kamaraj College of Engineering and Technology, Virudhunagar, Tamil Nadu, India
[6] Department of Information Technology, Kamaraj College of Engineering and Technology, Virudhunagar, Tamil Nadu, India
E-mail: [1] jagaenator@gmail.com, [2] madhusudhanece@anurag.edu.in, [3] karuna.mtech1976@gmail.com, [4] sknpotnuri@gmail.com, [5] ramyaads@kamarajengg.edu.in, [6] vakaimalarit@kamarajengg.edu.in

ABSTRACT

Food security planning accurate forecasts of crop yields are critical. Yet, combining heterogeneous agricultural information and modelling complex spatial and temporal dependencies are still great challenges. The current study suggests a Context-Aware Adaptive Spatio-Temporal Graph Transformer (CA-ASTGT) that enhances crop-yield prediction by adaptively combining multi-source information, integrating spatio-temporal information, and fusing features with context. The framework is based on a mixture of source-specific encoders and learned source weighting, a graph-based spatial message passing, a temporal Transformer encoder, and a context-conditioned fusion gate. A regression head predicts crop yield, and a relative yield-shortfall indicator is used in the production-risk assessment. The model is tested against the CY-Bench regional benchmark and is field-level YieldSAT independent test. The comparison is based on the root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and normalised RMSE (nRMSE) with the performance of ConvLSTM, spatio-temporal graph convolutional networks (STGCN), and a spatio-temporal graph Transformer (STGT). When compared with baselines, CA-ASTGT performs best on CY-Bench with an RMSE of 0.61 t/ha, an MAE of 0.44 t/ha, an R^2 of 0.89, and an nRMSE of 13.7%. On YieldSAT, it has an RMSE of 0.72 t/ha and an R^2 of 0.84. Ablation results show that all major components are contributing to predictive performance. The results here show the potential of the CA-ASTGT for crop-yield forecasting and for screening for production shortfalls, but more testing is needed to determine the generalisability and operational efficiency of the CA-ASTGT.

Keywords: *Agricultural Yield Forecasting; Spatio-Temporal Graph Transformer; Multi-Source Data Integration; Context-Aware Feature Fusion; Food Security; Production-Risk Assessment*.

1. INTRODUCTION

The need for governments, traders, and farming communities to estimate yields before harvest time to organize relief measures, pricing, trade, and storage led to the development of crop-yield estimation [1], [2]. Initial practices were based on field surveys, expert opinion and statistical regression of past yields on seasonal weather, providing useful national information with limited sensitivity to local variation. This has led to a

compounding cost of inaccuracy in forecasts due to climate variability, more extreme events, and growing demand, especially when food supply is very reliant on domestic production. With the development of satellite remote sensing, gridded weather products, and harmonised agricultural statistics, it is now possible to replace coarse and late estimates with timely and spatial explicit forecasts [3], [4]. The research problem is then based on the increasing gap between the data that is available today and the narrow capacity of traditional approaches to manage this data for food security planning.

The use of machine learning has been increasing and has been gradually superseding the statistical yield models. Deep architectures like Gaussian-process models based on remote-sensing histograms, and convolutional and recurrent networks, directly learn the nonlinear crop–environment relationship from the data; convolutional recurrent units and spatio-temporal graph networks are used to model dynamics that change across space and time. Additionally, graph neural networks have been integrated with recurrent layers to use the geographical neighborhood information of the yield observations and Transformer attention provides a flexible solution for long-range temporal dependence. Newly developed benchmarks (CY-Bench for sub-national forecasting, YieldSAT for field-level prediction of high resolution) allow for consistent evaluation [5], [6]. Although weather, soil, imagery, and crop records have varying levels of resolution, completeness, and reliability, and the usefulness of each of these sources varies with location and crop growth stage, integration is a key challenge [7]. Models which treat all inputs the same, or which have inputs with fixed importance, thus risk discarding useful information from reliable inputs, while picking up noise from unreliable ones, and this limit only worsens as more data streams are added.

The research problem is defined as: How to effectively integrate different kinds of agricultural data, capture the complex spatial and temporal interactions and enhance the agricultural yield forecasting ability to provide reliable and timely food security decision-making? In formal terms, the problem can be formulated as follows: Given a set of agricultural regions with a spatial graph of relations among them, each represented by a number of source-specific feature sequences and some contextual variables, the problem is to learn a mapping to the yield of each region at the forecast time, and from the learned mapping, an indicator of the potential production shortfall. The mapping should be maintained even if some of the data sources are missing or noisy, should reflect the spatial structure of the data and should also reflect the temporal one, and should not leak test information when generalising from regional statistics to field observations.

Better yield estimates directly impact farmers, policymakers and agricultural planners. Farmers can adapt their input choice, marketing and insurance activities, planners can optimize storage and procurement and logistics resources, and food security agencies can detect production gaps in time to take action. In today's environment of climate variability and shifting commodity markets, accurate forecasts that adjust to the quality of the data and are tied to risk information are valuable. The proposed framework captures the relationships between weather, soil, crop growth, and geography and has been evaluated using both regional and field-level data to determine if the benefits can be extended from the local to the regional level of decision making. The work highlights outputs that can be read as a risk message, since the former is increasingly consumed by non-specialists, and hence, the improvements to forecasting makes a difference to their everyday realities in terms of agricultural and food security governance.

The novelty of this work does not lie in any one of the four mechanisms per se but in the combination of the four, which is a single trainable framework. First, source contributions are learned adaptively for each location and time, and with unavailable sources masked, so that no missing observations are used as evidence. Second, the integrated representation is used in combination with Second, graph-based spatial learning and Transformer-based temporal attention. Third, a context-conditioned gate allows changing the balance between learned spatio-temporal information and contextual information according to the environmental conditions, crop identity and growth stage, rather than relying on a pre-fixed spatio-temporal fusion rule. Fourthly, forecasts are converted to a relative yield-shortfall indicator and the design is validated by ablation of components as well as by independently testing it on a second dataset of different spatial resolution. Combined, they take steps toward moving beyond accuracy-oriented forecasting and toward an adaptive to data quality, explicit spatial and temporal structure, and directly usable for production-risk screening approach.

In response to the gaps identified in the literature, the study pursues four objectives that are addressed

in turn by the integration, spatio-temporal, fusion, and assessment components of the proposed model:

1. To develop an adaptive multi-source data integration mechanism that effectively combines heterogeneous agricultural data to improve agricultural yield forecasting.
2. To develop a spatio-temporal graph Transformer model that jointly captures spatial relationships and temporal dependencies for accurate agricultural yield forecasting.
3. To design a context-aware feature-fusion mechanism that dynamically learns the importance and interactions of multi-source agricultural features under varying environmental and crop-growth conditions.
4. To evaluate the proposed framework's ability to identify potential agricultural production shortfalls and support timely, data-driven food security assessment and decision-making.

The crop-yield forecasting is enhanced by context-aware feature fusion, combined spatial-temporal modelling, and adaptive multi-source integration is examined. In comparison to baseline models, equal source weighting, and direct feature concatenation, it is hypothesised that these elements lower prediction errors.

2. RELATED WORK

2.1 Overview of Relevant Literature

Past production was regressed on climatic variables in early yield models, which were more interpretable, but formed the assumption of largely linear and homogeneous relationships. These assumptions were loosened by machine-learning methods, and even more by deep learning: Gaussian-process layers over satellite-derived histograms produced county-level estimates and CNN–RNN hybrids demonstrated a combined use of spatial and sequential learning to make predictions of yield from raw remote-sensing inputs [8], [9]. Spatio-temporal forecasting research also led to the development of spatio-temporal cells, such as convolutional recurrent units and graph convolutional networks, which propagate information across a spatial topology, but also take into account temporal dynamics [10], [11]. Models based on graphs and recurrent layers for agriculture have demonstrated that the estimation of yields is improved by the geographic context, which is achieved by the neighbourhood aggregation. This is now possible with the availability of CY-Bench and YieldSAT, where methods can be compared on harmonised regional data or high-resolution multimodal field data. Despite these progresses, most of the studies assess one data scale, are based on one dominant input type, or report on accuracy without considering how predictions are used in food security decision making, allowing for frameworks that consider integration, structure, context and application all together.

2.2 Key Theories and Relevant Concepts

The approach proposed is based on several concepts. In agriculture regions, where the condition is spatially autocorrelated, graph convolution and attention-based aggregation are formalizations that are more appropriate for how a node updates its representation based on neighbours. Self-attention considers all positions of a sequence simultaneously, giving the model the freedom to emphasise key phenological events, without the sequential bottleneck of recurrent networks [12], [13], [14]. The treatment of each data source as a modality that is encoded in a shared latent space and the flexibility of gated or attention-based fusion, which allows the contribution of each modality to fluctuate based on the input, are examples of multimodal learning. Lastly, food security is typically defined by four components: availability, access, utilisation and food security stability; yield projections primarily relate to food security availability, and thus the production gap from a reference yield is a proxy indicator of risks that do not fully capture food security [15]. The accuracy of such a proxy indicator will depend on forecasts not only being correct on average but also good at the bottom of the yield distribution where production concerns are relevant.

2.3 Gaps in the Literature

All the reviewed studies revealed good improvements in prediction accuracy, but four limitations emerged during their review which justify the present study:

1. **Limited adaptive integration of multi-source data.** Sources with different availability, quality and relevance are typically concatenated or statically weighted, limiting the forecasting effectiveness when sources vary.
2. **Inadequate modelling of spatial and temporal dependencies.** Spatial relationships or temporal dynamics can be represented well in many approaches, but only a few capture both at the same time in changing agricultural conditions.

3. **Limited context-aware feature fusion.** The fixed feature importance assumes that no other changes existed in the environmental conditions and crop growth stages and the regional characteristics.
4. **Insufficient connection between forecasting and food security.** Forecast accuracy is highlighted over forecasting to action on shortfalls and regional risk.

3. METHODOLOGY

3.1 Proposed Framework Overview

The paper introduces a novel model named Context-Aware Adaptive Spatio-Temporal Graph Transformer (CA-ASTGT) for forecasting yield and forecasting food security. The framework is designed in five phases: Multi-source Data, Adaptive Source Integration, Spatio-Temporal Graph Representation Learning, Context Aware Feature Fusion, and Yield Prediction and Shortfall Assessment, as shown in **Figure 1**. The four research objectives are covered by these stages: integration is covered by the first, the graph is covered by the second, the context gate is covered by the third, and the shortfall indicator is covered by the fourth. CY-Bench facilitates model development and initial evaluation while YieldSAT offers an independent test for generalisation at a finer spatial resolution.

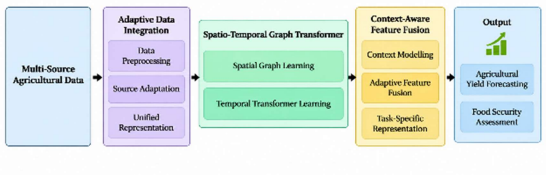


Figure 1. Overall architecture of the proposed CA-ASTGT framework.

3.2 Dataset Description

Two complementary datasets are used, and their characteristics are summarised in Table 1. CY-Bench is the primary source of regional crop-yield records with environmental predictors, while YieldSAT supplies field-level yield measurements with multimodal inputs for secondary validation. Because they differ in spatial resolution, crop coverage, and feature availability, the datasets are processed independently, with feature definitions and target units harmonised where appropriate. Invalid records and samples without a valid yield target are removed, continuous missing values receive median imputation, and missingness indicators are retained where informative. Continuous features are standardised with statistics estimated only on training data, as in Eq. (1), where x_j is the original feature, μ_j^{train} and σ_j^{train} are its training-set mean and standard deviation, and ϵ prevents division by zero. Partitioning is chronological and region-aware to limit leakage, and YieldSAT is reserved exclusively for secondary evaluation. Using earlier years and training regions for model creation and held-out years and locations for final testing, CY-Bench was divided geographically and chronologically into 70% training, 15% validation, and 15% testing data.

$$x_j^* = \frac{x_j - \mu_j^{\text{train}}}{\sigma_j^{\text{train}} + \epsilon} \quad (1)$$

Table 1. Dataset characteristics and preprocessing.

Characteristic	CY-Bench	YieldSAT
Purpose	Regional yield forecasting	Field-level yield validation
Crops	Maize and wheat	Corn, soybean, wheat, rapeseed
Coverage	25+ countries, six continents	Argentina, Brazil, Uruguay, Germany
Spatial resolution	Administrative regions	10 m × 10 m yield samples
Temporal coverage	Predictors 2003–2023; yield years vary	2016–2024
Input sources	Yield statistics, weather, soil, remote sensing	Yield, Sentinel-2, weather, soil, topography
Data partition	70/15/15 train/validation/test, chronological and region-aware	Independent geographic validation after filtering

3.3 Adaptive Multi-Source Data Integration

Let the input at region i and time t consist of M source-specific feature groups $X_{i,t}^{(m)}$ covering meteorology, soil, satellite-derived features, and historical yield. Each group is mapped to a common latent representation $h_{i,t}^{(m)} = f_m\left(X_{i,t}^{(m)}\right)$ by a learnable source encoder f_m. A scoring function g_m then estimates the relative contribution of each source from its encoding and the agricultural context $c_{i,t}$, and the scores are normalised by a softmax as in Eq. (2). The integrated representation is the weighted sum in Eq. (3), so source contributions can differ across locations and periods instead of being fixed. Unavailable sources are masked before normalisation so that missing observations are never treated as valid information.

$$\alpha_{i,t}^m = \frac{\exp\left(g_m\left(h_{i,t}^m, c_{i,t}\right)\right)}{\sum_{r=1}^{M} \exp\left(g_r\left(h_{i,t}^r, c_{i,t}\right)\right)} \quad (2)$$

$$z_{i,t} = \sum_{m=1}^{M} \alpha_{i,t}^{m} \, h_{i,t}^{m}$$

3.4 Spatio-Temporal Graph Transformer

Integrated representations are organised on an agricultural graph $G = (V, E, A)$, in which nodes are administrative regions or field units and edges follow geographic proximity and, where justified, feature similarity. The graph is built from training data and predefined geography only, and a k-nearest-neighbour configuration is the initial design, examined through sensitivity analysis. Spatial message passing in Eq. (4) combines each node with its neighbourhood $\mathcal{N}_i$ using normalised edge weights a_{ij}, learnable transformations W_s and W_n, and a nonlinearity ϕ. The resulting spatial representations of the last L observations are passed to a temporal Transformer encoder, Eq. (5), whose positional encoding preserves temporal order and whose self-attention selects the observations most relevant to the forecast.

$$s_{i,t} = \phi\left(W_s z_{i,t} + \sum_{j \in \mathcal{N}_i} a_{ij} \, W_n z_{j,t}\right) \quad (4)$$

$$T_i = \text{TransformerEncoder}\left(s_{i,t-L+1}, \ldots, s_{i,t}\right) \quad (5)$$

3.5 Context-Aware Feature Fusion

The spatio-temporal representation is combined with contextual variables such as environmental conditions, crop identity, and growth stage where the dataset supports them. A context encoder produces $q_{i,t} = f_c(c_{i,t})$, which drives the gate in Eq. (6), where σ is the sigmoid function, $\|$ denotes concatenation, and W_β b_β are learnable. The fused representation in Eq. (7) interpolates element-wise $\odot$ between learned spatio-temporal information and projected context through W_c. The gate therefore balances the two information streams dynamically, which addresses the third objective by letting feature contributions change with agricultural conditions.

$$\beta_{i,t} = \sigma\left(W_\beta \left[T_i \,\|\, q_{i,t}\right] + b_\beta\right) \quad (6)$$

$$u_{i,t} = \beta_{i,t} \odot T_i + \left(1 - \beta_{i,t}\right) \odot W_c q_{i,t} \quad (7)$$

3.6 Yield Prediction and Food Security Assessment

A linear regression head maps the fused representation to the predicted yield in Eq. (8), expressed in tonnes per hectare after inverse transformation when targets are standardised, and the network is trained by minimising the mean squared error in Eq. (9) over N training samples. To connect forecasting with food security, the prediction is compared with a reference yield derived from historical information or an agricultural benchmark, never from test outcomes. The relative shortfall in Eq. (10) measures the proportional deficit with respect to this reference, and larger values signal greater potential production concern. The indicator is a proxy for production risk, not a complete food security measure, since broader conclusions require information on production volumes, availability, access, and vulnerability.

$$\hat{y}_{i,t} = W_o u_{i,t} + b_o \quad (8)$$

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 \quad (9)$$

$$R_i = \frac{\max\left(0, y_i^{\text{ref}} - \hat{y}_i\right)}{y_i^{\text{ref}} + \epsilon} \quad (10)$$

3.7 Implementation and Training

The model is optimised with AdamW using the mean squared error objective, validation RMSE selects the checkpoint, and early stopping limits overfitting; the configuration is listed in Table 2, and the graph components rely on PyTorch Geometric. ConvLSTM, STGCN, and STGT baselines share data partitions, target definitions, and metrics. Ablations remove adaptive source integration, graph-based spatial learning, temporal Transformer learning, and context-aware fusion one at a time. Performance is reported by RMSE, MAE, R^2, and nRMSE over repeated seeds, and paired Wilcoxon signed-rank tests are applied to matched results where pairing is justified. The complete procedure is given in Algorithm 1.

Table 2. Hyperparameters and hardware/software configuration.

Parameter	Setting	Parameter	Setting
Nearest neighbours k	5	Optimiser / loss	AdamW / MSE
Graph layers / hidden dim.	2 / 128	Learning rate / decay	10^{-4} / 10^{-5}
Transformer layers / heads	3 / 4	Batch size / epochs	32 / 100
Window length L	5 observations	Early-stopping patience	15 epochs
Fusion mechanism	Context-conditioned gate	Dropout / seeds	0.20 / 5
Software	Python 3.11, PyTorch 2.4	Hardware	RTX 4090 (24 GB), 64 GB RAM

Algorithm 1. CA-ASTGT training procedure.

1. Load and preprocess CY-Bench; reserve YieldSAT for secondary evaluation.
2. Fit imputers and normalisation parameters on training data only.
3. Encode available sources and compute masked adaptive source weights.
4. Construct the agricultural graph and compute spatial node representations.
5. Apply temporal Transformer attention and context-aware gated fusion.
6. Predict yield and minimise MSE with AdamW; select the best checkpoint by validation RMSE with early stopping.
7. Evaluate on the held-out CY-Bench test set and on YieldSAT; run baselines, ablations, and statistical tests.
8. Compute yield-shortfall indicators using a predefined reference yield.

4. RESULTS

4.1 Experimental Setup

Experiments were conducted using a 70/15/15 training-validation-test split on CY-Bench, with YieldSAT reserved for secondary validation. Performance was evaluated using RMSE, MAE, R^2 and nRMSE, representing prediction error, explained variance and normalised error. Checkpoint selection was based on validation RMSE, and all models employed the same goal definitions and data splits. The test set was only used for the final check. To evaluate cross-dataset transfer, inputs were mapped to the trained feature schema for YieldSAT without parameter re-estimation.

4.2 Overall Prediction Performance

Figure 2 shows a comparison between the predicted and observed yields for the held-out CY-Bench test set. The points are concentrated along the y=x line, suggesting good agreement in both low and high yield conditions. There was some mild regression toward the mean at higher yields, as evidenced by increasing dispersion with increasing yield. The results show that the model is fairly predictive, but it is not enough to rely on the visual agreement only to verify the reliability of the model. The residuals have a mean of zero and some larger errors may be due to unusual seasons, such as droughts or very good growing seasons. The errors demonstrate the need for more studying production-shortfall detection and uncertainty-aware forecasting for food security applications.

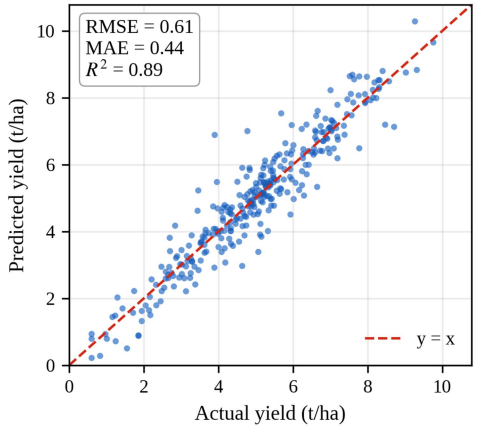


Figure 2. Actual versus predicted yield on held-out CY-Bench test observations.

4.3 Comparative Analysis

Table 3 shows a comparison between CA-ASTGT and three baseline models, which have been used in similar experiments under the same experimental conditions. CA-ASTGT consistently outperforms other methods across all evaluation metrics in acquiring the lowest RMSE, MAE, nRMSE and the highest R^2. The progression from ConvLSTM to STGCN and STGT highlights the potential benefits of graph-based spatial learning and temporal attention. Adaptive source weighting and context-aware fusion are seen to have an impact on predictive performance, further improvements over STGT. The slight decreases in RMSE and MAE suggest improvements in the overall error and mean prediction error, respectively, with the increase in R^2 signifying increased explained variance. The results confirm the effectiveness of the spatial and temporal modelling approach combined with adaptive multi-source integration, but validation at different agricultural settings is required to be sure of generalisability.

Table 3. Baseline model comparison on the CY-Bench test set.

Model	RMSE (t/ha)	MAE (t/ha)	R^2	nRMSE (%)
ConvLSTM	0.82	0.61	0.78	18.4
STGCN	0.75	0.55	0.82	16.8
STGT	0.69	0.50	0.85	15.5
CA-ASTGT	**0.61**	**0.44**	**0.89**	**13.7**

4.4 Ablation Study

Table 4 evaluates the contribution of each CA-ASTGT component by removing one component at a time. The complete model achieves the best performance, with an RMSE of 0.61t/ha, an MAE of 0.44t/ha, an $(R^2)$of 0.89, and an nRMSE of 13.7%. After taking out the graph-based spatial learning and adaptive source integration parts, the

performance drops the most when the temporal Transformer is taken out. The degradation decreases when context-aware feature fusion is removed, indicating that this element further enhances the learnt representations.

Equal source weighting, an identity graph, temporal pooling, or direct feature concatenation, respectively, were used to replace the eliminated component in each version while maintaining the identical data partitions and training process. The results show that contextually aware feature fusion, adaptive source integration, spatial learning, and temporal modelling all help CA-ASTGT perform better in its predictions.

Table 4. Component-wise ablation analysis on the CY-Bench test set.

Model configuration	RMSE (t/ha)	MAE (t/ha)	R^2	nRMSE (%)
Without adaptive source integration	0.70	0.51	0.85	15.7
Without graph-based spatial learning	0.73	0.53	0.83	16.4
Without temporal Transformer	0.76	0.56	0.82	17.0
Without context-aware feature fusion	0.67	0.48	0.86	15.0
STGT baseline	0.69	0.50	0.85	15.5
Complete CA-ASTGT	0.61	0.44	0.89	13.7

4.5 Generalization and Robustness

CA-ASTGT is compared in **Figure 3** to the performance on the CY-Bench test set and the independent YieldSAT dataset. The model has a strong R2 of 0.84, and despite the increase in errors on YieldSAT, there is potential for generalisation across datasets. These differences in performances could be due to difference in spatial resolution, difference in crop types, difference in regional condition and difference in measurement variability. Normalised error offers a means of interpreting the result of comparing the results of different datasets whose yield distributions vary. Evaluation of the model, without the estimation of the parameters, provides evidence of transferability, but differences in characteristics of the datasets make it difficult to make direct comparisons. Further testing in other crops, climates and seasons is required to confirm the robustness and generalisability of the CA-ASTGT in other agroecosystems.

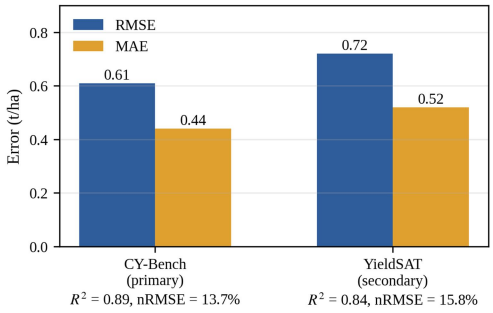


Figure 3. Cross-dataset performance on CY-Bench and YieldSAT.

4.6 Statistical Significance and Computational Efficiency

The statistical comparisons and computational costs of CA-ASTGT and the baseline models are summarised in Table 5. The performance differences between CA-ASTGT and the baseline models were assessed using the paired Wilcoxon signed-rank test, with a significance level of 0.05. The test was applied to paired results obtained under the same evaluation protocol. Inference latency is reported to be within a few milliseconds per sample, and CA-ASTGT's adaptive integration and context-aware fusion modules contribute to both its parameter count and inference time.

From the deployment point of view, these results provide an understanding of the trade-off between accuracy and computation. However, the importance of inference latency is not the only consideration because the agricultural yield forecasting is generally used for periodic planning. So, the benefit of the extra complexity must be judged in conjunction with the gains in RMSE, MAE, $R^2$and nRMSE achieved.

Table 5. Statistical significance and computational efficiency.

Model	Parameters (M)	Inference (ms/sample)	Paired test vs. CA-ASTGT
ConvLSTM	2.10	4.8	$p < 0.01$
STGCN	1.65	3.6	$p < 0.01$
STGT	2.35	5.2	$p < 0.05$
CA-ASTGT	**2.80**	**6.1**	Reference

5. DISCUSSION

5.1 Interpretation of the Main Findings

The results show that there is no single module that contributes to the increased accuracy, but rather, it comes from the interplay of the proposed components. This capability of adaptive weighting gives the model more weight to data sources that

are informative, and less weight to data sources that are noisy or missing, which is crucial in operational environments with data of varying quality. Graph propagation provides information from neighboring areas with similar climate and management conditions, while temporal focus focuses on growth periods that affect final yield. The context gate then fine-tunes the strength of the influence of the context information on the representation. The temporal and spatial structure is the ground for the forecasting, while the adaptive integration and contextual fusion are enhancements to boost the reliability under heterogeneous environment conditions, which is aligned with the first three objectives. This is possible agronomically because yield is a summation of the weather over a season, soil water holding capacity and neighbourhood-level management, and representation that either spatial or temporal dimension is missing will lose structure that the target will depend on. To achieve the 4th objective, the value of the shortfall indicator is based on the quality of the predictions in the lower end of the yield distribution, where accurate and well-calibrated predictions are most important.

5.2 Comparison with Existing Approaches

In comparison to convolutional recurrent models, the proposed framework has more natural modeling of irregular regional topology and long-range temporal dependence, while compared to spatio-temporal graph convolutional networks [16], [17] it eliminates the local temporal filtering by replacing it with self-attention [18]. In previous agricultural studies with deep Gaussian processes, CNN-RNN hybrids, or graph–recurrent hybrids [19], [20] have shown the added value of learned features and geographic context, with the major drawback being that they typically assumed a static subset of inputs without considering the conditions or crop stage. The present design overcomes this drawback without compromising the spatio-temporal benefits of these models. As different datasets, crops and protocols are used in published studies, direct numerical comparison is not appropriate and the best evidence of the relative merits is the controlled baseline comparison. The margin is particularly informative here, due to the shared spatio-temporal backbone with STGT: the margin shows integration and fusion processes, not a different backbone.

5.3 Practical Implications

The framework provides regional forecasts which can be revised in real time as more weather and satellite data becomes available, assisting practitioners as they decide on the input to use, inputs to purchase, inputs to store, and how to trade such inputs well before harvest. The relative shortfall indicator is a straightforward screening tool to identify areas that might require attention, especially at sub-national level, to reduce the risk of a national average concealing localised supply failures that impact on vulnerable households, thereby enabling planners to prioritise resources to verify and provide relief where the estimated shortfall from the reference is greatest. High level of robustness to missing sources allows deployment in regions where monitoring is sparse, and moderate computational requirements enable frequent retraining and large-scale inference on standard GPUs. The benefits rely on the careful selection of the reference yield, and on the incorporation of the forecasts with other food availability and access information, which means that the outputs should complement and not supplant the expert judgement. Organisations that follow the approach should also track the quality of inputs over time as sensor changes or reporting methods may change the reliability of the source.

5.4 Limitations and Future Research Directions

The following restrictions should be noted. The shortfall indicator is only a proxy for production risk, and a complete assessment of food security should include information on stocks, trade, prices and household vulnerability. The indicator also relies on a reference yield, which means that the definition of this reference yield is important and choices of imputation and graph should be reconsidered when using the framework in areas not included in the study area. The graph is built statically from proximity and similarity, and does not necessarily reflect dynamic interactions like trade or shared irrigation; and the evaluation is limited to a certain number of crops and regions. Assumptions are also made when establishing mapping features between regional and field level data, which can have a transfer impact. Future research could include the use of probabilistic forecasts including the incorporation of calibrated uncertainty, dynamic or learnable graph structures, interpretability analysis of source weights and gates, other food security indicators and testing in extreme climate scenarios and other regions and crops.

6. CONCLUSION

To boost agricultural yield forecasting, this study proposed the Context-Aware Adaptive Spatio-Temporal Graph Transformer (CA-ASTGT) that leverages adaptive multi-source data

integration, simultaneous spatial and temporal dependency modelling, and context-aware feature fusion and production-shortfall assessment. The experimental results showed that the performance of CA-ASTGT outperformed ConvLSTM, STGCN, and STGT on the CY-Bench data set, with an RMSE of 0.61 t/ha, an MAE of 0.44 t/ha, an R^2 of 0.89 and an nRMSE of 13.7%. Further evaluation on YieldSAT showed an RMSE of 0.72 t/ha and R^2 of 0.84 which may allow for cross-dataset generalisation. The contributions of the major components of the model were confirmed by ablation experiments. The study provides an integrated forecasting framework that links the predictive modelling to production-shortfall screening in agricultural planning. Uncertainty-aware forecasting, dynamic graph construction, interpretable feature fusion, and extensive validation of crop types, regions and seasons will be explored in future studies, enhancing the reliability and applicability of the model for food security decision-making.

REFERENCES

[1] S. Mahalakshmi, A. Jose Anand, and P. Partheeban, "Soil and crop interaction analysis for yield prediction with satellite imagery and deep learning techniques for the coastal regions," *J. Environ. Manage.*, vol. 380, p. 125095, Apr. 2025, doi: 10.1016/j.jenvman.2025.125095.

[2] Z. Liu *et al.*, "In-season crop yield prediction: State of the art and future research direction," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 146, p. 105129, Feb. 2026, doi: 10.1016/j.jag.2026.105129.

[3] P. Muruganantham, S. Wibowo, S. Grandhi, N. H. Samrat, and N. Islam, "A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing," *Remote Sens.*, vol. 14, no. 9, p. 1990, Apr. 2022, doi: 10.3390/rs14091990.

[4] Md. A. Jabed and M. A. Azmi Murad, "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability," *Heliyon*, vol. 10, no. 24, p. e40836, Dec. 2024, doi: 10.1016/j.heliyon.2024.e40836.

[5] S. M. Shawon, F. B. Ema, A. K. Mahi, F. L. Niha, and H. T. Zubair, "Crop yield prediction using machine learning: An extensive and systematic literature review," *Smart Agric. Technol.*, vol. 10, p. 100718, Mar. 2025, doi: 10.1016/j.atech.2024.100718.

[6] S. Sarkar *et al.*, "Crop Yield Prediction Using Multimodal Meta-Transformer and Temporal Graph Neural Networks," *IEEE Trans. AgriFood Electron.*, vol. 2, no. 2, pp. 545–553, Sep. 2024, doi: 10.1109/TAFE.2024.3438330.

[7] D. Radočaj, M. Jurišić, I. Plaščak, and L. Galić, "Satellite Remote Sensing for Crop Yield Prediction: A Review," *Agriculture*, vol. 16, no. 4, p. 417, Feb. 2026, doi: 10.3390/agriculture16040417.

[8] S. Dangi, S. K. Mullapudi, C. S. Raghaw, S. S. Dar, M. Z. U. Rehman, and N. Kumar, "A multi-temporal multi-spectral attention-augmented deep convolution neural network with contrastive learning for crop yield prediction," *Comput. Electron. Agric.*, vol. 239, p. 110895, Dec. 2025, doi: 10.1016/j.compag.2025.110895.

[9] K. Meghraoui, T. Racharak, K. Ait El Kadi, S. Bensiali, and I. Sebari, "A new integrated neurosymbolic approach for crop-yield prediction using environmental data and satellite imagery at field scale," *Artif. Intell. Geosci.*, vol. 6, no. 1, p. 100125, Jun. 2025, doi: 10.1016/j.aiig.2025.100125.

[10] M. Beyer, R. Ahmad, B. Yang, and P. Rodríguez-Bocca, "Deep spatial-temporal graph modeling for efficient NDVI forecasting," *Smart Agric. Technol.*, vol. 4, p. 100172, Aug. 2023, doi: 10.1016/j.atech.2023.100172.

[11] T. Gong, X. Zhu, Y. Chen, and S. Xiong, "Temporal-Aware Spatial Interaction Transformer for Crop Yield Prediction Based on Multisensor Satellite Image Time Series," *IEEE Geosci. Remote Sens. Lett.*, vol. 22, pp. 1–5, 2025, doi: 10.1109/LGRS.2025.3627552.

[12] M. Kallenberg *et al.*, "CY-Bench: a comprehensive benchmark dataset for sub-national crop yield forecasting," *Earth Syst. Sci. Data*, vol. 18, no. 6, pp. 3997–4018, Jun. 2026, doi: 10.5194/essd-18-3997-2026.

[13] D. Paudel *et al.*, "CY-Bench: A comprehensive benchmark dataset for subnational crop yield forecasting." AgML (https://www.agml.org/), Mar. 11, 2025. doi: 10.5281/ZENODO.15008945.

[14] A. Joshi, B. Pradhan, S. Chakraborty, R. Varatharajoo, S. Gite, and A. Alamri, "Deep-Transfer-Learning Strategies for Crop Yield Prediction Using Climate Records and

Satellite Image Time-Series Data," *Remote Sens.*, vol. 16, no. 24, p. 4804, Dec. 2024, doi: 10.3390/rs16244804.

[15] F. Mena *et al.*, "Adaptive fusion of multi-modal remote sensing data for optimal sub-field crop yield prediction," *Remote Sens. Environ.*, vol. 318, p. 114547, Mar. 2025, doi: 10.1016/j.rse.2024.114547.

[16] P. Patro, R. Azhagumurugan, R. Sathya, K. Kumar, T. R. Kumar, and M. V. S. Babu, "A hybrid approach estimates the real-time health state of a bearing by accelerated degradation tests, Machine learning," in *2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Bengaluru, India: IEEE, Dec. 2021, pp. 1–9. doi: 10.1109/ICSTCEE54422.2021.9708591.

[17] S. Mahmood, R. Hasan, and S. Ahmad, "HSE-GNN-CP: Spatiotemporal Teleconnection Modeling and Conformalized Uncertainty Quantification for Global Crop Yield Forecasting," *Information*, vol. 17, no. 2, p. 141, Feb. 2026, doi: 10.3390/info17020141.

[18] M. Qiao *et al.*, "KSTAGE: A knowledge-guided spatial-temporal attention graph learning network for crop yield prediction," *Inf. Sci.*, vol. 619, pp. 19–37, Jan. 2023, doi: 10.1016/j.ins.2022.10.112.

[19] F. Huber, A. Inderka, and V. Steinhage, "Leveraging Remote Sensing Data for Yield Prediction with Deep Transfer Learning," *Sensors*, vol. 24, no. 3, p. 770, Jan. 2024, doi: 10.3390/s24030770.

[20] G. Xiao *et al.*, "Progress and Perspectives of Crop Yield Forecasting With Remote Sensing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 13, no. 3, pp. 338–368, Sep. 2025, doi: 10.1109/MGRS.2025.3571906.