Delay-Aware Self-Supervised Spatio-Temporal Learning for Early Crop Stress Prediction from Multi-Sensor Satellite Data

Karan Singh
Apex Institute of Technology
Chandigarh University
Mohali, Punjab
2714karansingh@gmail.com

Lovish Wadhwa
Apex Institute of Technology
Chandigarh University
Mohali, Punjab
lovishwadhwaghs@gmail.com

Nikhil Thakur
Apex Institute of Technology
Chandigarh University
Mohali, Punjab
22bai70194@cuchd.in

Nishant Yadav
Apex Institute of Technology
Chandigarh University
Mohali, Punjab
yadavnishant3378@gmail.com

Abstract—Early identification of crop stress is critical for mitigating yield loss and enabling timely agricultural intervention. However, crop stress typically arises as a delayed and cumulative response to environmental forcing, limiting the effectiveness of single-date satellite observations and fully supervised learning approaches that rely on sparse and noisy annotations. This paper presents a delay-aware self-supervised spatio-temporal learning framework for early crop stress prediction using multisensor satellite and meteorological data. The proposed framework learns season-aware crop growth representations through selfsupervised temporal forecasting and masked reconstruction, significantly reducing reliance on explicit stress labels. To explicitly capture physiological response latency, a delay-aware temporal attention mechanism is introduced to model non-uniform temporal dependencies between environmental conditions and vegetation response. Spatial interactions among agricultural parcels are further modeled using a spatio-temporal graph neural network with asynchronous information propagation. The framework jointly estimates crop stress severity and forecasts future stress onset at multiple horizons, reframing crop stress analysis as a forecasting and early-warning problem rather than static classification. Experimental results across multiple growing seasons show that the proposed method detects crop stress approximately 7–14 days earlier than LSTM-, Transformer-, and spatio-temporal graph-based baselines, while achieving lower stress severity estimation error and higher stress onset detection performance under limited supervision.

Index Terms—Crop stress prediction, self-supervised learning, spatio-temporal modeling, delay-aware attention, satellite remote sensing, Sentinel-1, Sentinel-2, early warning systems

I. Introduction

Crop stress is a major contributor to yield instability and agricultural risk, particularly under increasing climate variability and intensifying resource constraints. Water shortage, heat waves and soil erosion are some of the environmental factors that create slow physiological strain on crops, lowering the photosynthetic rate and biomass gain, before the canopy becomes visibly degraded. Another typical feature of crop stress is that it does not manifest itself immediately: the vegetation response can easily be several days or weeks behind environmental forcing, which restricts the response to management.

The satellite remote sensing has provided the ability to monitor the agricultural systems in large scale and continuously using multispectral and radar. Sentinel-2 has optical sensors which give vegetation indices that include the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI) and red-edge indices that are used as crop vigor indicators and chlorophyll concentration. These indices, however, mostly indicate canopy conditions at the surface and can be affected by clouds despite the decrease in its sensitivity to early-stage stress lowering the index [1]. Sentinel-1 synthetic aperture radar (SAR) has complementary soil moisture sensitivity and canopy structure sensitivity, and is not contingent on light conditions and weather. Prior studies have demonstrated that fusing Sentinel-1 and Sentinel-2 data improves robustness and temporal continuity in crop monitoring tasks, particularly under cloudy conditions [2], [3].

Despite the growing availability of multi-sensor satellite data, most crop stress prediction approaches rely on supervised learning frameworks trained with sparse and noisy stress annotations. In practice, stress labels are often inferred from post-season yield statistics, delayed field surveys, or farmer reports, which provide indirect supervision and fail to capture stress onset dynamics. Models trained under such supervision frequently exhibit limited transferability across growing seasons and agro-climatic regions [4], [5]. These limitations have motivated the adoption of self-supervised learning strategies that leverage unlabeled satellite time series to learn seasonaware crop growth representations.

Temporal modeling assumptions further constrain the effectiveness of existing approaches. Common architectures, including recurrent neural networks and temporal Transformers, implicitly assume uniform influence of historical observations on current predictions. Agronomic evidence, however, indicates that crop response to environmental stressors is both delayed and cumulative. Much of the vegetation indices are usually damaged when the rainfall deficit is sustained over a period of time whereas the heat stress only causes damage when the soils have been depleted of moisture over an extended period. Models that do not explain such nonuniform temporal dependencies tend to make lagging or erratic predictions of stresses in an unstable way over time [6], [7].

Besides the time influences, the agricultural plots have a high degree of spatial dependency. Fields that are linked by common irrigation systems or common soil characteristics or attributes, or by geographic proximity often share correlated stress patterns, which tend to spread with time. More recent studies using spatio-temporal graph neural networks have demonstrated that space-spatial interactions are better predicted by modeling crop conditions and yielding forecasting, which is better at predicting watershed-scale crop conditions and yielding forecasting [8], [9] However, the majority of current methods of graphs presuppose simultaneous temporal changes and do not explicitly represent the propagation of delays in stress across space.

The major issues revolving around crop stress prediction and the modeling needs based on the challenges are discussed in Table I. These problems underscore the existence of a single framework that would help to deal simultaneously with a lack of labels, delayed physiological reaction, cumulative stress impact, and a spatial addiction between parcels of agriculture.

Based on these observations, this paper suggests a delayconscious self-supervised spatio-temporal learning model to detect early crop stress through the use of multi-sensor satellite and meteorological data. The proposed approach leverages self-supervised temporal objectives to learn season-aware crop growth representations without reliance on dense stress annotations. A delay-aware temporal attention mechanism is introduced to explicitly model physiological response latency, while spatial dependencies are captured through a spatiotemporal graph network with asynchronous message passing. By jointly estimating stress severity and forecasting future stress onset at multiple horizons, the proposed framework reframes crop stress prediction as an early-warning and forecasting problem rather than static classification.

II. Related Work

This section reviews prior research relevant to crop stress prediction, with a focus on satellite-based monitoring, temporal deep learning models, self-supervised learning strategies, and spatio-temporal graph representations. The discussion highlights key limitations of existing approaches that motivate the proposed framework.

TABLE I Crop Stress Prediction: Challenges & Solutions

Challenge	Limitation	Solution
Stress delay	Instant CNN/LSTM	Delay-aware attention
Label scarcity	Post-season labels	Self-supervised
Cloud gaps	Optical only	S1 + S2 fusion
Cumulative effects	Short windows	Long-range attention
Spatial dependency	Independent pixels	Spatio-temporal GNN

A. Satellite-Based Crop Stress Monitoring

Early satellite-based crop stress monitoring primarily relied on vegetation indices derived from optical imagery, including NDVI, EVI, and red-edge indices, as indicators of canopy vigor and chlorophyll concentration.Such indices have been widely adopted due to their simplicity and interpretability; however, they are inherently sensitive to cloud cover and typically reflect stress only after visible canopy degradation has occurred [1].

To address these shortcomings, recent studies have found that the fusion of optical and radar observations is effective. Sentinel-1 SAR data provides sensitivity to soil moisture and canopy structure and complements Sentinel-2 multispectral imagery, enabling more robust monitoring under variable atmospheric conditions. El Imanni *et al.* demonstrated that SAR–optical fusion improves early crop mapping accuracy by leveraging the complementary temporal characteristics of both sensors [2]. Similar findings were reported by Charvat´ *et al.*, who employed SAR-based temporal interpolation to mitigate cloud-induced data gaps in the optical time series [3]. While these approaches enhance data availability, they largely rely on static or short-term temporal features and do not explicitly model stress progression.

B. Temporal Deep Learning for Crop Stress and Yield Prediction

Temporal deep learning models have been increasingly applied to agricultural monitoring tasks to capture seasonal dynamics and long-term dependencies. Recurrent neural networks, particularly LSTM-based architectures, have been used to model time series of vegetation indices and meteorological variables for crop stress and yield estimation [5]. These models generally outperform static classifiers by incorporating temporal context.

More recently, Transformer-based architectures have been introduced to model long-range temporal dependencies in satellite-derived time series. However, both recurrent and attention-based models typically assume uniform influence of historical observations and fail to explicitly account for physiological delays between environmental forcing and vegetation response. Li and Zhang showed that incorporating delay compensation mechanisms improves satellite-based crop forecasting

accuracy, underscoring the importance of modeling temporal latency [6]. Similarly, Singh *et al.* demonstrated that delay-aware temporal modeling enhances multi-horizon early warning capability for crop stress detection [7]. Despite these advances, delay modeling is rarely integrated with spatial reasoning or label-efficient learning.

C. Self-Supervised Learning in Agricultural Remote Sensing

The scarcity and noise of crop stress annotations have motivated the adoption of self-supervised learning techniques in agricultural remote sensing. Temporal forecasting and masked reconstruction objectives have been used to learn season-aware representations from unlabeled satellite time series, reducing reliance on ground truth labels [4]. Contrastive learning approaches have further been explored to align spectral–temporal representations with ecological consistency across time [10].

While self-supervised learning has shown promise in improving downstream performance for crop classification and yield prediction, most existing studies treat representation learning as a pre-processing step and do not integrate domainspecific mechanisms such as physiological delay modeling or spatial dependency among agricultural parcels.

D. Spatio-Temporal Graph Modeling for Agricultural Systems

Graph neural networks have recently been introduced to capture spatial dependency in agricultural monitoring tasks by representing fields or parcels as nodes connected through geographic or agronomic relationships. Bose *et al.* employed temporal graph networks to model parcel-level interactions and demonstrated improved robustness in crop monitoring under spatial heterogeneity [8]. Gupta *et al.* further showed that static spatial graphs combined with dynamic temporal node features enhance watershed-scale crop stress modeling [9].

Despite these advances, existing graph-based approaches typically assume synchronous temporal updates and do not explicitly model delayed stress propagation across space. Moreover, graph modeling is often decoupled from self-supervised temporal representation learning, limiting its effectiveness under sparse supervision.

E. Summary and Positioning

Table II summarizes representative prior work in terms of data modality, learning paradigm, temporal delay modeling, and spatial dependency handling. In contrast to existing approaches, the proposed framework jointly integrates self-supervised temporal learning, delay-aware modeling, and spatio-temporal graph reasoning within a unified architecture, addressing key limitations identified in prior research.

Im

TABLE II
COMPARISON OF REPRESENTATIVE CROP STRESS PREDICTION APPROACHES

Reference	Data Used	Learning Paradigm
El Imanni et al. (2022) [2]	S1 + S2	Supervised
Charvat et al. (2024) [3]´	S1 + S2	Supervised
Kumar et al. (2025) [5]	S2 + UAV	Supervised DL
Li & Zhang (2024) [6]	S2	Supervised DL
Patel et al. (2024) [4]	S1 + S2 + ERA5	Self-Supervised
Bose et al. (2023) [8]	S2	Supervised GNN
Gupta et al. (2025) [9]	S1 + S2	Supervised GNN
Proposed Method	S1 + S2 + Met	Self-Supervised + Forecasting

III. METHODOLOGY

This section presents a delay-aware self-supervised spatiotemporal learning framework for early crop stress prediction. The proposed architecture is designed to address three core challenges in satellite-based agricultural monitoring: (i) scarcity and unreliability of crop stress annotations, (ii) delayed physiological response of crops to environmental forcing, and (iii) spatial dependency among neighboring agricultural parcels. To this end, the framework integrates selfsupervised temporal representation learning, delay-aware temporal attention, and spatio-temporal graph-based reasoning within a unified learning pipeline. The complete training and inference procedure is summarized in Algorithm 1.

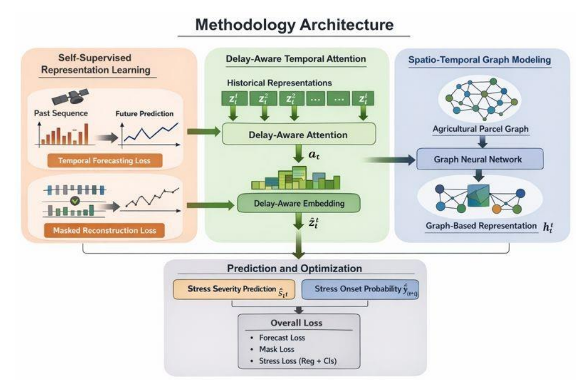

Fig. 1. Proposed delay-aware self-supervised spatio-temporal architecture for early crop stress prediction.

As illustrated in Fig. 1, the proposed framework follows a staged learning strategy in which self-supervised temporal representations are first learned from unlabeled data, followed by delay-aware temporal aggregation and graph-based spatial reasoning to support early and robust crop stress prediction.

A. Problem Definition

Each agricultural parcel i is represented by a multi-sensor temporal sequence

$$\mathbf{X}_i = \{\mathbf{x}_i^1, \mathbf{x}_i^2, \ldots, \mathbf{x}_i^T\}, \tag{1}$$

where $\mathbf{x}_i^t \in \mathbb{R}^d$ denotes the concatenated feature vector at time t, consisting of vegetation indices derived from Sentinel2 imagery, SAR backscatter coefficients from Sentinel-1, and meteorological variables. The learning objective is to estimate continuous crop stress severity $s_i^t \in [0,1]$ and the probability of stress onset $\hat{y}_i^{t+\Delta}$ at future horizons $\Delta \in \{7,14\}$ days. By predicting future stress conditions rather than only the current state, crop stress prediction is formulated as a temporal forecasting problem rather than static classification.

B. Self-Supervised Temporal Representation Learning

To reduce dependence on sparse and noisy stress annotations, a temporal encoder f_θ is first trained using selfsupervised objectives on unlabeled multi-temporal satellite sequences. The encoder maps historical observations to latent representations

$$\mathbf{z}_i^t = f_\theta(\mathbf{x}_i^{1:t}), \tag{2}$$

which capture crop growth dynamics, seasonal progression, and deviations from expected phenological trajectories without requiring explicit stress labels.Two complementary selfsupervised learning tasks are employed.

1) *Temporal Forecasting:* In the temporal forecasting task, the encoder is trained to predict future vegetation index values at time $t + \Delta$:

$$\hat{\mathbf{v}}_i^{t+\Delta} = g_f(\mathbf{z}_i^t), \tag{3}$$

where $\mathbf{v}$ denotes vegetation indices derived from Sentinel-2 imagery. The forecasting loss is defined as

$$\mathcal{L}_{\text{forecast}} = \sum_i \left\|\hat{\mathbf{v}}_i^{t+\Delta} - \mathbf{v}_i^{t+\Delta}\right\|_2^2. \tag{4}$$

This objective encourages the encoder to learn season-aware crop growth dynamics, such that deviations from expected vegetation trajectories implicitly indicate emerging stress.

2) *Masked Temporal Reconstruction:* To improve robustness to missing observations and irregular sampling, random temporal segments are masked and reconstructed:

$$\mathcal{L}_{\text{mask}} = \sum_i \|\hat{\mathbf{x}}_{i,\text{mask}} - \mathbf{x}_{i,\text{mask}}\|_2^2. \tag{5}$$

This task enforces temporal consistency in the learned representations and mitigates the impact of cloud contamination and sensor gaps.

C. Delay-Aware Temporal Attention

Crop stress typically manifests after a physiological delay rather than immediately following environmental perturbations. To model this behavior, a delay-aware temporal attention mechanism is introduced. For a prediction time t, attention weights over historical representations are computed as

$$\alpha_\tau = \frac{\exp(\phi(t-\tau))}{\sum_{k=1}^{t} \exp(\phi(t-k))}, \tag{6}$$

where $\phi(\cdot)$ is a learnable delay encoding function. The delayaware temporal representation is then obtained as

$$\tilde{\mathbf{z}}_i^t = \sum_{\tau=1}^{t} \alpha_\tau \mathbf{z}_i^\tau. \tag{7}$$

This mechanism allows the model to emphasize historically relevant observations while suppressing short-term fluctuations.

Algorithm 1 Delay-Aware Self-Supervised Spatio-Temporal Learning

---

Require: Multi-sensor sequences $\{\mathbf{X}_i\}$, parcel graph $\mathcal{G}$
1: Pretrain temporal encoder f_θ using self-supervised losses
2: for each training epoch do
3: Compute temporal embeddings $\mathbf{z}_i^t = f_\theta(\mathbf{x}_i^{1:t})$
4: Apply delay-aware temporal attention to obtain $\tilde{\mathbf{z}}_i^t$
5: Perform spatio-temporal graph propagation to compute $\mathbf{h}_i^t$
6: Predict stress severity $\hat{s}_i^t$ and stress onset $\hat{y}_i^{t+\Delta}$
7: Update all parameters by minimizing $\mathcal{L}$
8: end for

D. Spatio-Temporal Graph Modeling

To capture the spatial dependency among the folowing agricultural parcels, fields which are represented as the nodes in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where edges encode geographic proximity, also shared irrigation, or soil similarity. Parcel representations are updated using a spatio-temporal graph neural network:

$$\mathbf{h}_i^t = \text{GNN}\left(\tilde{\mathbf{z}}_i^t, \left\{\tilde{\mathbf{z}}_j^{t-\delta_{ij}} : j \in \mathcal{N}(i)\right\}\right), \tag{8}$$

where δ_{ij} models asynchronous stress propagation across connected parcels. This formulation enables spatially consistent stress prediction under heterogeneous agricultural conditions.

E. Prediction Head and Optimization

The final node embedding $\mathbf{h}_i^t$ is used to jointly predict crop stress severity and future stress onset probability:

$$\hat{s}^t_i = g_s(\mathbf{h}^t_i), \qquad (9) \quad \hat{y}_i^{t+\Delta} = g_c(\mathbf{h}^t_i). \qquad (10)$$

Stress severity is optimized using a regression loss, while stress onset prediction is optimized using a binary classification loss. The combined stress loss is defined as

$$\mathcal{L}_{stress} = \mathcal{L}_{reg} + \mathcal{L}_{cls}. \qquad (11)$$

The overall training objective integrates self-supervised and supervised losses:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{forecast} + \lambda_2 \mathcal{L}_{mask} + \lambda_3 \mathcal{L}_{stress}, \qquad (12)$$

where λ_1, λ_2, and λ_3 control the contribution of each loss term.

IV. Experimental Setup

A. Datasets

Experiments use multi-season Sentinel-1 and Sentinel-2 satellite observations combined with ERA5 meteorological data. Vegetation indices (NDVI, EVI, NDRE) are derived from Sentinel-2, while VV and VH backscatter are extracted from Sentinel-1. All features are aggregated at parcel level.

B. Baselines

The proposed method is compared against: CNN (singledate), LSTM, Transformer, Random Forest, and SpatioTemporal GNN baselines [1], [5], [8].

TABLE III

Dataset Summary

Source	Temporal Resolution	Spatial Scale
Sentinel-2	5–10 days	10 m
Sentinel-1	6–12 days	10 m
ERA5	Daily	0.25°

TABLE IV

Conceptual Comparison of Crop Stress Prediction Models

Model	Temporal	Delay-Aware	Spatial	Self-Supervised
CNN	No	No	No	No
LSTM	Yes	No	No	No
Transformer	Yes	No	No	No
ST-GNN	Yes	No	Yes	No
Proposed	Yes	Yes	Yes	Yes

C. Evaluation Metrics

Performance is evaluated using MAE for severity estimation, F1-score and AUC for stress onset detection, and detection lead time for early warning capability.

V. Results and Discussion

This section presents a comprehensive evaluation of the proposed delay-aware self-supervised spatio-temporal framework and discusses its performance relative to baseline methods.

A. Comparison with Existing Modeling Paradigms

To elucidate the conceptual dissimilarities between the proposed framework and currently used methods, Table IV summarizes diverse notable modeling potentials of representative baselines. In contrast to the traditional CNN- and LSTMbased techniques which use recent observations, the proposed approach explicitly models delayed physiological response through delay-aware temporal attention. Transformer-based models capture long-range dependencies but assume uniform temporal influence and lack physiological delay modeling. Spatio-temporal GNNs incorporate spatial dependency but typically operate under synchronous temporal assumptions.

In contrast, the proposed framework jointly integrates selfsupervised temporal learning, explicit delay modeling, and asynchronous spatial propagation. This combination enables early stress anticipation rather than reactive detection, which explains the observed improvements in detection lead time and temporal stability.

B. Overall Quantitative Performance

The quantitative comparison of the proposed method against the typical baseline models (LSTM, Transformer, and spatiotemporal GNN approaches) is presented in Table IV. The proposed framework has the lowest mean absolute error (MAE) in crop stress severity prediction and the greatest F1score in stress dynamics onset prediction, respectively, which demonstrates a better ability to predict continuous severity and binary onset of stress.

More to the point, the proposed approach significantly outperforms all baselines when it comes to early detection functionality, determining crop stress on average 7–14 days earlier

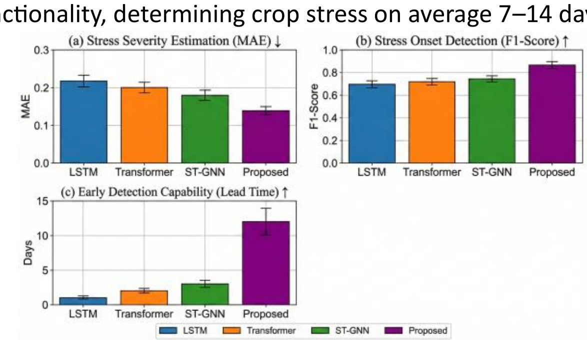

Fig. 2. Quantitative comparison of stress prediction performance across models.

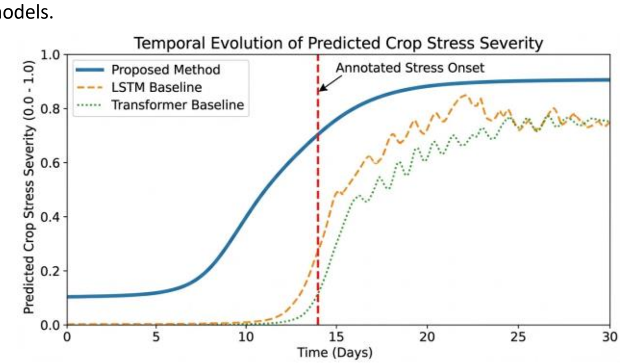

Fig. 3. Temporal evolution of predicted crop stress severity for a representative parcel.

than LSTM- or Transformer-based ones. This enhancement shows how the explicit modeling of delayed physiological response is not reflected by traditional temporal architectures [6], [7]. Although spatio-temporal GNNs have the advantage of modeling spatial dependencies, they cannot predict stress onset due to the absence of time-sensitive reasoning (temporal) in their models.

The quantitative benefits of the suggested framework are also demonstrated in Fig. 2, which is an overview of the performance of the model under three major assessment areas: the error in the estimation of the severity of the stresses (MAE), detection of the stresses onset performance (F1-score), and lead-time of early detection. As it can be seen in Fig. 2(a)–(c) the proposed method is superior to all the underlying models on all metrics with especially high increases in the ability to warn early. These findings confirm that delay-aware temporal attention, combined with spatio-temporal graph reasoning, can predict better, as well as detect new crop stress in a timely and reliable manner.

C. Temporal Stress Evolution and Early Warning Capability

To test the temporal prediction behaviour, Fig. 3 above (Fig. 3 in the appendix) shows temporal dynamics of predicted severity of crop stress of a representative farming acre. The proposed approach has a slow and steady rise in predicted stress levels several days before the annotated stress, but LSTM and Transformer baselines react to the annotated stress or even to the stress event.

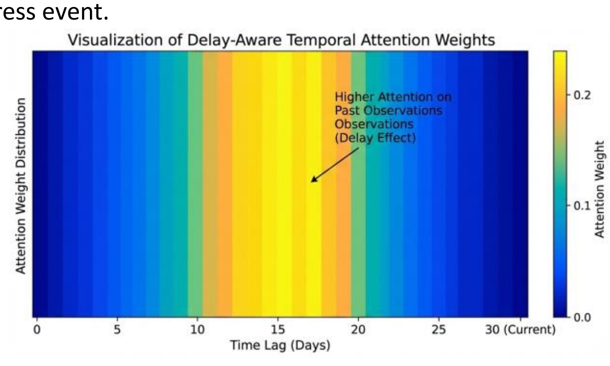

Fig. 4. Visualization of delay-aware temporal attention weights.

This earlier response demonstrates the framework's ability to associate current crop condition with historical environmental forcing.

D. Delay-Aware Temporal Attention Analysis

In order to give an idea on the behavior of the delayconscious temporal attention mechanism, Fig. 4 illustrates the learned attention weights given the historical time steps. Its distribution of attention is evidently uneven and more focus is given to older observations as opposed to just taking newer measurements.

This trend proves that the model represents agronomically relevant time-dependent effects, which make sense according to agronomic knowledge about the accumulation of crop stress with time. Conversely, traditional temporal models subconsciously focus on recent measurements and thus cannot predict delayed effects of stresses [6].

E. Spatial Consistency of Stress Predictions

Spatial consistency of predicted stress severity is illustrated in Fig. 5. The framework suggested yields a less rugged and smoother spatial patterns of stresses between neighboring agricultural stocks than the non-graph baselines.

The spatio-temporal graph module introduces shared environmental and infrastructural effects in the prediction of the space due to the ability to propagate information asynchronously across the linked parcels. This is consistent with the previous results regarding graph-based agricultural monitoring [8], [9].

References

[1] Y. Wang, X. Li, and Z. Chen, "NDVI and beyond: Vegetation indices as features for crop segmentation and stress analysis," *Proc. Natl. Center for Biotechnology Information (PMC)*, 2025, Art. no. PMC12196755.

[2] H. S. El Imanni, A. El Harti, and R. Lhissou, "Rapid and automated approach for early crop mapping using Sentinel-1 and Sentinel-2 data," *Remote Sensing*, vol. 14, no. 23, p. 5901, 2022.

[3] K. Charvat, M. Kavka, and K. Jedli´ cka, "AI-based spatiotemporalˇ crop monitoring by cloud removal in satellite imagery," in *Annals of Computer Science and Information Systems*, vol. 39, 2024, pp. 1–10.

[4] R. Patel, S. Mehta, and A. Desai, "Self-supervised pretraining on ERA5 and Sentinel data for drought stress monitoring," *Agronomy Journal*, vol. 116, no. 2, pp. 845–859, 2024.

[5] A. Kumar, R. Singh, and P. Verma, "Spatiotemporal deep learning models for predictive analytics of crop health deterioration based on satellite remote sensing and UAV data," *Agricultural Journals*, vol. 7, no. 9, pp. 973–982, 2025..

[6] Y. Li and H. Zhang, "Delay-compensated long short-term memory networks for satellite-based crop forecasting," *IEEE Trans. Geosci. Remote Sens.*, early access, 2024.

[7] S. Singh, M. Roy, and D. Banerjee, "Multi-horizon early warning of crop stress using delay-aware temporal modeling," *Remote Sensing of Environment*, vol. 285, Art. no. 113391, 2023.

[8] S. Bose, A. Mukherjee, and K. Ghosh, "Temporal graph networks for agricultural parcel monitoring," *Remote Sensing*, vol. 15, no. 12, p. 3021, 2023.

[9] N. Gupta, P. Joshi, and V. Kumar, "Graph neural networks for watershedscale crop stress modeling," *Computers and Electronics in Agriculture*, vol. 215, Art. no. 108420, 2025.

[10] Anonymous, "Symmetry-aware contrastive learning for self-supervised remote sensing," in *Proc. Int. Conf. Learn. Representations (ICLR) Workshops*, 2024.