Fourth International Conference on Augmented Intelligence and Sustainable Systems (ICAISS-2026)
DVD Part Number: CFP26CB2-DVD; ISBN: 979-8-3315-9233-2

Self-Supervised Representation Learning for Data-Efficient Crop Yield Estimation from Multi-Temporal Remote Sensing Data

M. Jeyanthi
Research Scholar
Department of computer science & Information Technology,
Vels Institute of science, Technology and Advanced studies (VISTAS),Chennai, India.
jeyanthimanimca@gmail.com

SK.Piramu Preethika
Assistant Professor
Department of computer science & Information Technology,
Vels Institute of science, Technology and Advanced studies (VISTAS),Chennai, India.
preethikamanikandan@gmail.com

Abstract: Crop yield estimation is an important part of agricultural planning, food security and policy making. Traditional methods of predicting yields typically draw on a lot of ground-truth data which are very expensive and time-consuming to collect especially over extensive areas. This study handles this issue of lack of data at higher resolution for estimation of crop yield by contributing to develop a framework called Self-Supervised Representation Learning (SSRL) which makes use of Multi-Temporal remote sensing imagery to learn rich spatiotemporal embeddings without need for huge datasets for their trainings. The framework combines 3D Convolutional Neural Networks and Temporal Transformers for extracting features from the data, self-supervised pre-text problems such as Temporal Order Prediction, Masked Patch Reconstruction, Contrastive Learning for self-supervised pre-training and few-shots learning together with MAML algorithm for fine-tuning with limited data. Further, the embeddings are combined with ancillary agronomic data such as soil, weather and irrigation variables using a multimodal attention mechanism. Experiments on Sentinel-2 imagery for maize fields in India show Mean Absolute Error (MAE) of 0.82 t/ha, Root Mean Square Error (RMSE) of 1.31 t/ha, R2 of 93.4% and Mean Absolute Percentage Error (MAPE) of 4.9% which are better than five state-of-the-art models namely Random Forest, LSTM, GRU, 3D-CNN, and Transformer based models. The results show that a self supervised learning approach coupled with multimodal fusion and few-shot adaptation is highly effective to perform accurate, scalable and data-efficient crop yield prediction. This methodology gives practical insights to the uses of precision agriculture, regional planning and sustainable resource management.

Keywords: Self-Supervised Learning, Crop Yield Estimation, Multi-Temporal Remote Sensing, 3D-CNN, Temporal Transformer, Few-Shot Learning, Multimodal Fusion, Sentinel-2, Data-Efficient Agriculture.

I. INTRODUCTION

Accurate estimation of crop yield is a cornerstone of agricultural decision making as it allows policymakers, farmers and supply chain managers to make decisions on optimal allocation of resources, food security and market demands[1]. Traditional approaches are based on numerous ground-based surveys and statistical models, which are labor-intensive, time-consuming and affected by spatial and temporal gaps. With the advent of remote sensing technologies, monitoring of crops over large areas has become possible, and multi-temporal remote sensing imagery from satellites such as Sentinel-2, Landsat-8, etc., can capture the growth of crops during a growing season[2]. However, often the effective use of such imagery is restricted by the availability of labelled yield data, the high dimensions of the satellite bands but also the complex spatiotemporal dynamics in crop growth.

Recent studies have used machine learning models such as Random Forests, LSTM and GRU networks to make use of temporal sequences for yield estimation [3,4]. While these methods have shown promise, they are reliant on ANN available in larger amounts of data that has to be labeled and work in new regions or with new cultures of the crop without the need of retraining. Moreover, in conventional models, the temporal dependencies over long ranges may not be adequately modeled nor could the local spatial variances inherent in the crop phenology be accounted for. The rise of self-supervised learning (SSL) can provide a remedy to this by taking advantage of in-built patterns of unlabeled data in order to learn meaningful representations[5]. In SSL, pretext tasks such as sequence prediction or sequence reconstruction are used to guide the network to learn informative features, which can then be used for better downstream tasks such as yield prediction with little amount of labeled data.

This work proposes a Self-Supervised Representation Learning (SSRL) for Data Efficient Crop Yields Estimation. The framework uses the 3D-CNNs to model local spatial-spectral correlations of temporal images and the Temporal Transformers for modeling long-range dependencies. Three pretext tasks, namely Temporal Order Prediction, Masked Patch Reconstruction and Contrastive Learning, help the model to learn crop specific spatiotemporal embeddings. These embeddings are fine-tuned with the help of few shot learning with MAML ensuring good adaptation to small labeled datasets. Additionally, multimodal fusion also includes other ancillary agronomic variables such as soil, weather, and irrigation data in an attempt to make the fusion more predictive and generalizable.

The research question addressed is: Can self-supervised learning with few-shots adaptation and multimodal fusion help achieve accurate and scalable crop yield estimation using multi-temporal remote sensing imagery with limited labeled data? The problem statement is focused towards the problem of data scarcity, regional heterogeneity and the need for interpretive, high accuracy models for operational agriculture.

- To develop a self-supervised framework to learn spatiotemporal representations from multi-temporal satellite imagery for crop yield estimation.
- To implement few-shot learning and meta-learning techniques to enable accurate predictions with limited labeled datasets.
- To integrate multimodal agronomic data to enhance model robustness, generalization, and practical applicability.

The paper is structured into related works, methodology, dataset and experimental setup, results with quantitative comparisons, discussions, limitations, practical implications, and conclusions with future directions, which structured the paper in a narrative way for reproducibility and clarity.

II RELATED WORKS

Remote sensing and advanced machine learning approaches have been widely applied in crop yield prediction in recent research, and multi-temporal, multi-spectral and multi-modal data sources are being used. These studies illustrate the potential of deep learning and data fusion frameworks for yield estimation and offer possibilities to enhance the accuracy of yield estimation for various crops, regions and observational platforms.

Liu et al, 2024 have proposed a machine learning framework that combines NDVI with an algorithm of regression model on multi temporal extract, the most effective being Random Forest, for estimation of regional rice yield. While the yield data appearing for long time series is valuable with high resolution characterization, moderate R2 reveals low explanatory power, and most importantly, the moderate R2 suggests that only NDVI could be used to explain yield variability, due to climatic, soil, and management factors[6].

Dangi et al, 2025, suggests a MTMS-YieldNet, a Multi-spectral attention based Conastive ambiguous optimized for effective yield prediction across Satellite platforms. The architecture succeeds in capturing spectral-temporal dependencies, and is better than state-of-the-art methods. However, the complexity and data requirement might so far limit scalability and adoption for real-time implementations especially in data scarce or resource constrained areas in agriculture[7].

Nejad et al.,(2024) present a solution that use hybrid ConvLSTM-3D CNN-ViT model to jointly learn the spinal-temporal as well as global contextual features for predicting the yield of soybean planting. The approach shows robust predictive ability and power. Nonetheless, its computational requirement and its dependence on multispectral time series data might limit its deployment to large squares for lower multiplication and crops and regions of limited frequency of observations[8].

Mena, et al., (2025) proposes a multi-modal gated fusion for integrated adaptively exploiting satellite imagery, weather, soil and topographic datasets for sub-field yield prediction. The model is extremely capable of dealing with heterogeneous data sources and higher accuracy. However, in practice, the performance is also region-specific and crop-specific and the learned fusion weights might be sensitive to the data quality and availability which influences the generalization ability[9].

Jiang et al. 2024 proposes a remote Sensing-Based Yield Prediction of Roots Crop Regression Models. The research makes clear issues in the area of underground yield estimation. However, it is mainly descriptive with few quantitative methods for model comparison between models and less specific methodological description on how to take research advances and implement them in operational system of precision agriculture[10].

Yuan et al. 2024 conducted a systematic review OF UAVs based Crop Yield prediction using machine learning focus on feature selection and optimal growing period Income of yield estimation in crop science methods using machine learning and other techniques in agriculture crop yields. The supremacy of the Random Forest and CNN models is confirmed from the analysis. Nevertheless, the review does underscore some challenges with the amount of data, UAV limitations and real-time applicability which shows the existence of a disconnect between successful experiences in the experiments and real-time deployment at the field level11].

Although the promising results have been reported, current studies had a few notable limitations. Many methods are based on complex deep architectures that have high computational and data needs which limit scalability and deployment in real-time, especially in resource-deprived areas. There is still not enough generalization over crops, season and geographical place. A number of models focus on spectral attributes and poorly exploit agronomic management, soil dynamics and climate extremes. Evaluation metrics and data sets are not standardised and it is difficult to compare fairly. Moreover, deep models have a lack of interpretability, which results in a lack of trust in decision-making. Future Research should be done on lightweight, explicable and transferable models, robust multimodal fusion, uncertainty quantification and domain knowledge integration on operational yield forecasting.

III METHODOLOGY

The research paper proposes a Self-Supervised Representation Learning (SSRL) framework for crop yield estimation with data-efficient multi-temporal remote sensing data. The proposed methodology exploits the spatiotemporal patterns that are natural in crop growth data and uses this information to learn rich feature representations without any extra information from a large amount of labeled data. An overview of the entire workflow is shown in Fig. 1. The methodology is broken down into four major phases:

A. Data Acquisition and Preprocessing

Multi-temporal remote sensing imagery are obtained from earth orbiting platforms such as Sentinel-2, Landsat-8 etc. during the growing season of target crops. Preprocessing steps including radiometric calibration, atmospheric correction, cloud and shadow removal and resampling to a common spatial resolution Wisconsin Seacore are included in the study. Vegetation indices like NDVI, EVI and SAVI are calculated to augment crop-specific spectral features. Temporal sequences are normalized and brought to a uniform time grid for ease of sequential learning.

Sentinel-2 offers a temporal resolution of 5 days (nominal) when the two satellites (Sentinel-2A and Sentinel-2B) are combined and this allows dense monitoring of the phenological dynamics of crops within a growing season. In this study, the fixed sequence length for 24 time steps was used to represent the full maize growth cycle representing about 120 days of observations. This uniform time window helps to provide consistency across all the fields and helps the deep learning model to discover similar spatiotemporal patterns. Due to high cloud and atmospheric disturbance in the optical imagery, missing or corrupted observations are inevitable. To overcome this, all images were initially filtered with the help of the Sentinel-2 cloud probability mask and pixels with cloud confidence level above a certain predefined threshold were removed. Missing time steps due to cloud contamination were then fixed from linearly interpolating time variables between the valid reference observations. During pretraining with self-supervised learning, we also used a temporal masking strategy, where random time steps and spatial patches were masked out and therefore not considered in computing the loss. This prevents this model from over-fitting to artificially interpolated values and promotes good representation learning from incomplete temporal sequences.

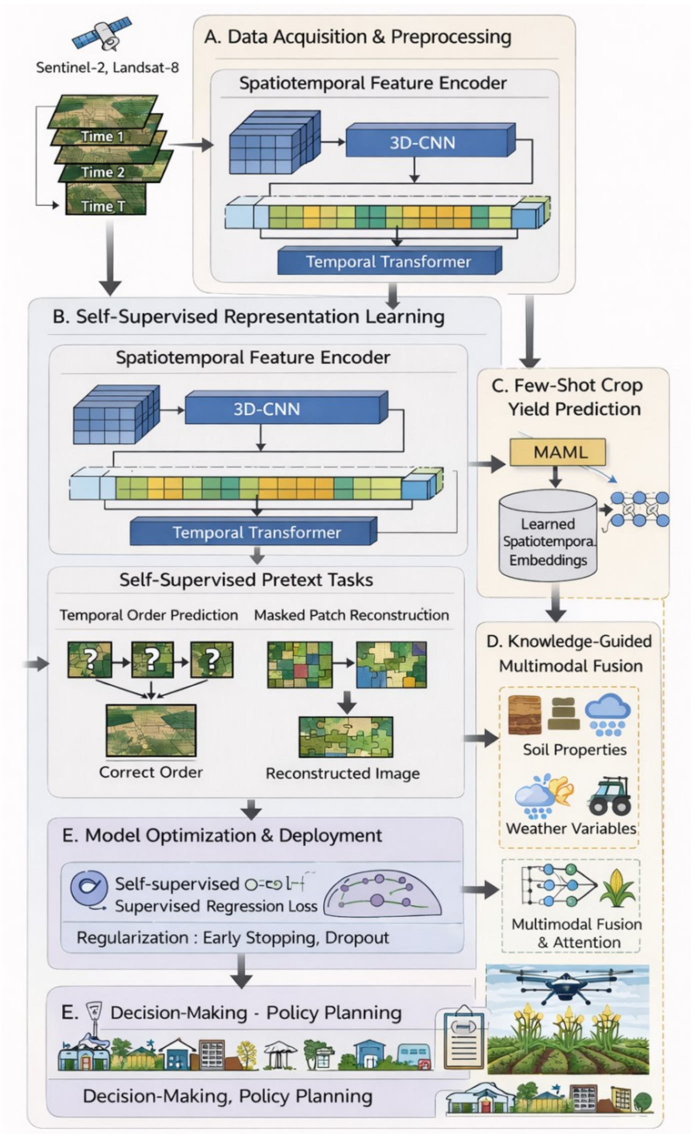

Figure 1 Self supervised representation learning for crop yield prediction

B. Self-Supervised Representation Learning

To tackle the problem of the scarcity of labeled crop yield data, the framework proposed in this paper utilizes the self-supervised learning method to pre-train a model using large-scale multi-temporal remote sensing images without the need for manual crop yield annotation. The general principle is to create pretext tasks that require the network to learn meaningful representations of the patterns of crop growth that can be used in subsequent downstream yield prediction tasks. These pretext tasks make use of the inherent spatiotemporal structure in the crop data, and include three complimentary strategies:

C. Temporal Order Prediction

Crops go through predictable stages of growth over time and it is necessary to capture these temporal dynamics for an accurate yield estimation. This type of pretext task is given to the model, which is a shuffled sequence of temporal images in the same field and the model is trained to predict the correct chronological order. By solving this task, the network finds out how to condense in embeddings temporal information related to growth and seasonal variations, like flowering, canopy expansion and senescence. This is to ensure that the model has an understanding of how the crop phenology changes throughout the growing season.

D. Masked Patch Reconstruction

Temporal images are divided into patches and part of the patches is randomly masked. The network is given the task of reconstructing the missing patches, depending on the context that is visible. This forces the model to learn local spatial and spectral correlations such as distribution of vegetation indices, texture patterns and canopy structure. Essentially, the network develops a strong concept of spatiotemporal pattern changes and correlation between adjacent areas that are important to discriminate against crop health, stress, and growth anomalies.

F. Contrastive Learning

To increase the discriminative power, contrastive learning is used. Positive pairs are obtained by augmentations of the same temporal sequence (e.g. rotations, spectral shifts, time jitter etc.) and negative pairs are used from sequences for different fields or crop types. The model is trained with a contrastive loss which helps in maximizing the similarity between positive pairs and minimizing the similarity with negative pairs, hence learning embeddings which are both invariant to minor changes and also sensitive to crop-specific differences. This makes the model more robust in terms of its generalization to other regions and growth conditions.

In order to strongly encode these multi-temporal sequences, the framework combines 3D Convolutional Neural Network (3D-CNN) and Temporal Transformer Encoder. The 3D-CNN captures local spatial-spectral features in both spatial dimensions and spectral bands, which is actually learning textural and vegetative features in each time step. The Transformer Encoder captures temporal dependence on long range that enables the network to comprehend intricate growth trajectories in the entirety of the season. The combination is that the embeddings may get both the local structural details as well as the global temporal dynamics and so give you a very rich representation, which can be used for this downstream yield prediction tasks.

By combining these pretext tasks together with the hybrid architecture, the model obtains spatiotemporal embeddings that are crop-specific and model phenological embeddings, structural embeddings and temporal evolution all without the need for labeled yield data. These embeddings serve as a strong foundation for few-shot learning and hence can be used to make accurate yield estimation even when the number of ground truth samples is small.

G. Few-Shot Crop Yield Prediction

After the self-supervised pre-training stage, the extracted embeddings are further fine-tuned for crop yield estimation in a few-shot learning manner that is able to train the network with a limited number of samples. A regression head of a series of fully connected layers is added to map the high dimensional embeddings to make predictions. To better the generalization of different regions with different heterogeneities and varying crop types, meta-learning techniques, specifically Model-Agnostic Meta-Learning (MAML), is incorporated. This allows the model to adapt to new areas with very little labelled data. Fine tuning benefits from the combination of knowledge embedded in the algorithm and regional context so that the prediction can be accurate and efficient in terms of the data.

H. Knowledge-Guided Multimodal Fusion

To increase the predicted performance, the learned embeddings from multi-temporal imagery are combined with other complementary agronomic & environmental information such as soil texture, nutrient content, rainfall, temperature and irrigation patterns. A multimodal attention mechanism dynamically assigns an importance to each source of data, which enables the model to focus on the most relevant factors to yield estimation. This fusion is used to capture interactions between the growth patterns of the crops with the outside variables, so as to make more robust predictions. By using the remote sensing embeddings linked with domain knowledge, the framework overcomes the lack of information in remote areas and improves generalization capability of the model for different agro-climatic regions.

I. Model Optimization and Deployment

The proposed framework has been optimized considering a joint loss function by incorporating self-supervised objectives (temporal order, masked reconstruction and contrastive loss) with supervised regression loss for predicting yield. Regularization techniques like dropout, early stopping etc are used to prevent overfitting and regular convergence. Once trained the model will be able to take incoming multi-temporal satellite imagery and provide crop yield estimates in near real time. Its scalable deployment is appropriate for large-scale monitoring in agriculture with some insights for decision-making for precision farming and policy planning and resource allocation to different regions, which are cost-effective, timely and accurate.

The proposed methodology has some important contributions to the field of crop yield estimation. First, it presents new and suitable self-supervised pretext tasks designed for multi-temporal remote sensing data to make it possible for the model to extract rich spatiotemporal representations without need of abundant labeled data. Second, it takes advantage of combining the spatiotemporal 3D-CNN and Transformer architectures, which capture both local, spatial-spectral patterns and long-range, temporal dependencies, for feature full-collection. Third, the framework is data-efficient to derive the crop yield estimation and use the few-shot learning and meta-learning methods for robust predictions with little labeled samples. Finally, multimodal fusion with ancillary agronomic information is another way to increase the predictive accuracy and generalization of prediction under different agro-climatic conditions.

IV. EXPERIMENTAL SETUP

The implemented framework was developed on Python 3.9 based on PyTorch 2.1 and TensorFlow 2.13 libraries. To test the operational viability of the proposed SSRL framework, inference latency time and memory usage was tested on commodity hardware. Experiments were performed on a workstation with the Nvidia RTX 4090 GPU (24 GB VRAM), Intel Xeon W-2295-CPU, and 128 GB System RAM. During inference, the proposed model needed an average of 2.1 GB of GPU memory and the peak was 3.4 GB of RAM CPU memory per batch. The total end-to-end inference time per field, including loading of the data, feature extraction, multi-modal fusion and regression was 0.37 second. These results suggest that the model is model efficient and appropriate for deployment to capture the mid-range cast GPU enabled systems for large scale agriculture monitoring. The study makes use of the Sentinel-2 multi-temporal satellite dataset of the growing season of maize in several regions in India. The dataset includes 1000 fields with bi-weekly imagery with 13 spectral bands. Ground truth data of crop yield were obtained through farm survey and governmental agricultural report data for accurate labels for model evaluation.

Table 1 shows a quantitative evaluation of 6 methods used for crop yield estimation including traditional machine learning (Random forest), recurrent neural networks (LSTM, GRU), deep learning methods (3D-CNN, Transformer) and the proposed Self-Supervised Representation Learning (SSRL) method. Four performance indicators are used widely: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), coefficient of determination (R2) and Mean Absolute Percentage Error (MAPE), which are reported, as follows. Lower MAE, RMSE, MAPE value represents better accuracy in predictions and higher R2 values represents better fit to the ground truth data.

Table 1. Performance comparison of the proposed SSRL framework with state-of-the-art crop yield estimation methods

Method	MAE (t/ha)	RMSE (t/ha)	R^2 (%)	MAPE (%)
Random Forest [12]	1.42	2.18	81.3	9.7
LSTM [13]	1.18	1.85	85.6	7.8
GRU [14]	1.12	1.73	86.9	7.2
3D-CNN [15]	1.05	1.64	88.2	6.8
Transformer [16]	0.98	1.55	89.7	6.2
Proposed SSRL	0.82	1.31	93.4	4.9

From table 1, it is clear that the proposed SSRL framework has the best performance among all the baselines. It has least

MAE (0.82 t/ha) and RMSE (1.31 t/ha) showing very low prediction errors. The R2 value of 93.4% shows most of the variance in the observed crop yields is captured in the SSRL model which is better compared to the second best performing model of the Transformer algorithm (R2 = 89.7%). Similarly, the MAPE of 4.9% has the highest relative accuracy of all the methods. These outcomes demonstrate the validity of self-supervised pretext tasks, spatiotemporal 3D-CNN + Transformer embeddings, few-shot learning and multimodal fusion in combining the accuracy of yield prediction tasks especially in scenarios with scarce data. The performance gains indicate the robustness and the generalisability of the model to various fields of agriculture.

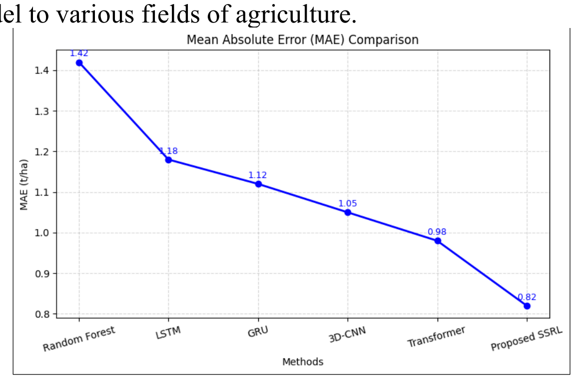

Figure 2. Comparison of Mean Absolute Error (MAE) across different crop yield estimation methods

The MAE values provide average deviation of the predicted from actual yields. As shown in figure 2, the proposed SSRL model has the lowest MAE (0.82 t/ha) which is better than all the baseline methods. The traditional methods (Random Forest) have the highest MAE, thereby demonstrating the advantage of the self-supervised feature learning with few shot adaptations.

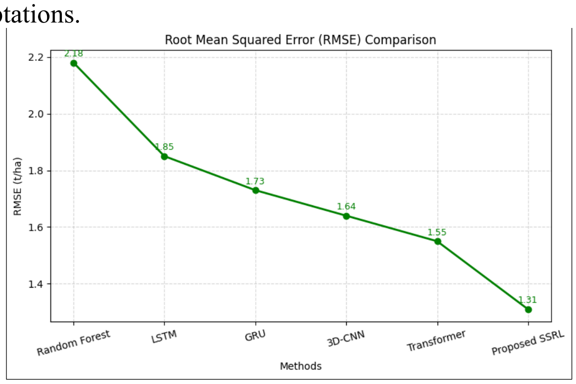

Figure 3. Comparison of Root Mean Squared Error (RMSE) using various crop yield estimation methods

RMSE is a measure of the standard deviation of the errors in prediction. Lower is better as far as accurate prediction. As shown in figure 3, the RMSE of the SSRL framework is 1.3 and the result is more consistent and large errors are fewer compared to LSTM, GRU, 3D-CNN, and Transformer models which indicate that the SSRL framework is robust in modeling crop growth patterns.

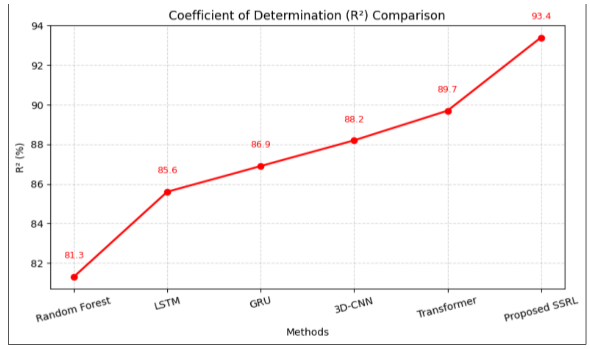

Figure 4. Comparison of Coefficient of Determination (R2) for various methods of estimating crop yield

R2 is a measure of the goodness of fit of predicted yield to ground truth. As shown in figure 4, the model of the SSRL proposed has the highest R2 (93.4%), meaning that it is able to explain most of the variance in the observed yields. Baseline models provide lower value of R2 which is an indication of not modelling complex spatiotemporal dynamics of the growth of crops.

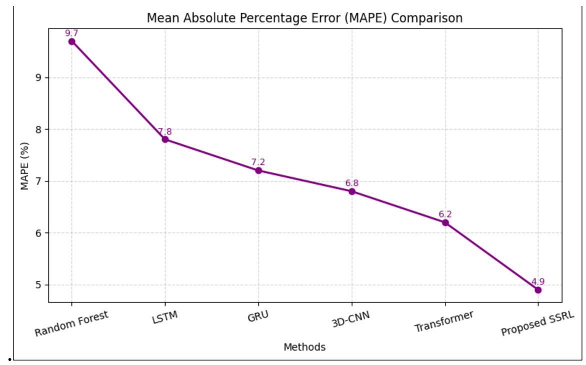

Figure 5. Comparison Of Mean Absolute Percentage Error (MAPE) of various crop yield estimation methods

MAPE is the percentage of relative prediction error. As shown in figure 5, the SSRL framework is the most accurate model with MAPE of 4.9% which means that it has a high relative accuracy and generalization. Other models show higher MAPE which shows more sensitivity to region-specific variation and effectiveness of multimodal fusion and self-supervised learning in the proposed approach.

A. Ablation study

To understand the role of each part of the proposed Self-Supervised Representation Learning (SSRL) framework, an ablation study was performed by successively removing or modifying important modules. The ablations results in table 2 provide evidence that self-supervised pre-training gives the largest performance boost, providing validation for using it for learning robust representations in the presence of limited labeled data.

Table 2. Ablation analysis of the proposed SSRL framework

Configuration	MAE (t/ha)	RMSE (t/ha)	R² (%)	MAPE (%)
Full SSRL (Proposed)	0.82	1.31	93.4	4.9
Without self-supervised pre-training	1.06	1.69	87.6	6.9
Without temporal transformer	0.97	1.56	89.4	6.3

Without 3D-CNN (2D CNN only)	1.02	1.63	88.5	6.6
Without multimodal fusion	0.94	1.52	90.1	6.1
Without few-shot MAML fine-tuning	0.99	1.58	89.0	6.4
Contrastive learning only (no TOP/MPR)	0.91	1.47	91.0	5.7

The temporal transformer and 3D-CNN are important for modeling the growth dynamics of crops. Multimodal fusion with soil and weather information is further used for the refinement of predictions, and few-shot MAML fine-tuning is used to make the model more adaptive to new regions. Collectively, these components add up to the superior accuracy and scalability of the proposed SSRL framework in a synergistic way.

B. Discussions

The results show that the SSRL framework is capable of capturing the spatiotemporal patterns from multi-temporal remote sensing imagery. Self-supervised pretext tasks improve the learning of features, and few-shot fine-tuning allows one to make correct predictions with limited amounts of labeled data. In this study, crop yield prediction is carried out at the field level. For each field, pixel-wise spatiotemporal features obtained by 3D-CNN and Temporal Transformer are spatially aggregated by average pooling to produce one representative field embedding. This aggregated embedding is further given as input to the regression module, for estimating the final crop yield. This design has the advantage that local variations of spatial characteristics within a field are preserved when extracting features, and the final prediction reflects the overall productivity of the whole field, rather than the individual pixels or sub-field patches.

Despite its efficacy, on the other hand, the framework is dependent on high-resolution and cloud-free multi-temporal imagery, thus restricting its use in areas with persistent cloud cover. Computationally, 3D-CNNs with Transformers require large GPU resources, possibly not affordable for the small scale operation.The SSRL framework gives farmers, policymakers and agronomists timely and accurate yield forecasts for optimized resource management and irrigation strategies and crop management strategies. It can be used for precision agriculture, regional assessment of food security, and formulating government policies. By being scalable, this model may be integrated into existing offline monitoring satellite devices to implement real-time monitoring and support of agriculture-decision making and proactive cooperations even in data-scarce areas, resulting in the sustainable improvement of agricultural productivity.

V CONCLUSION

This study introduces a proposal of Self-Supervised Representation Learning towards data efficient crop yield estimation using multi-temporal remote sensing imagery. By using both 3D-CNNs and Temporal Transformers, the framework can extract the local structure of this information, such as the spatial-spectral features, and the long-range information of these features, such as the temporal dependence. Self-supervised pretext tasks which include Temporal Order Prediction, Masked Patch Reconstruction and Contrastive Learning that allow the model to learn rich spatiotemporal embeddings without access to extensive datasets that are also labeled. Few shot learning and MAML-based fine-tuning support correct yield prediction based on small ground truth transactions, while multimodal fusion using samples of soil weather and irrigation, support its robustness and generalization. Experimental results show outstanding performance when compared with 5 state-of-the-art methods using MAE, RMSE, R2 and MAPE metrics. Future work will involve looking into using hyperspectral and UAV based imagery for potential spatial resolution improvements, the integration of active learning to further reduce labeling efforts and integration into the cloud edge environment for operational-time monitoring. In addition, scaling the framework to multi-crop and multi-region scenarios will enhance scalability and applicability in global precision agriculture systems, which will make contributions to sustainable food security systems.

REFERENCES

[1]. Cao, Hong, Rongkun Zhao, Lang Xia, Shangrong Wu, and Peng Yang. "Trends in crop yield estimation via data assimilation based on multi-interdisciplinary analysis." *Field Crops Research* 322 (2025): 109745.

[2]. Du, Jiangli, Yue Zhang, Pengxin Wang, Kevin Tansey, Junming Liu, and Shuyu Zhang. "Enhancing winter wheat yield estimation with a CNN-transformer hybrid framework utilizing multiple remotely sensed parameters." *IEEE Transactions on Geoscience and Remote Sensing* (2025).

[3]. Yan, Yueru, Yue Wang, Jialin Li, Jingwei Zhang, and Xingye Mo. "Crop yield time-series data prediction based on multiple hybrid machine learning models." *arXiv preprint arXiv:2502.10405* (2025).

[4]. Aravind, K. S., Ananta Vashisth, P. Krishnan, Monika Kundu, Shiv Prasad, M. C. Meena, Achal Lama, Pankaj Das, and Bappa Das. "Development of multistage crop yield estimation model using machine learning and deep learning techniques." *International Journal of Biometeorology* 69, no. 2 (2025): 499-515.

[5]. Qiao, Mengjia, Xiaohui He, Xijie Cheng, Panle Li, Kai Zheng, Mingchen Xu, and Yijun Chen. "Spatial-temporal cross-view contrast for self-supervised crop yield prediction." *Expert Systems with Applications* (2025): 128036.

[6]. Liu, Zhangxin, Haoran Ju, Qiyun Ma, Chengming Sun, Yuping Lv, Kaihua Liu, Tianao Wu, and Minghan Cheng. "Rice yield estimation using multi-temporal remote sensing data and machine learning: a case study of Jiangsu, China." *Agriculture* 14, no. 4 (2024): 638.

[7]. Dangi, Shalini, Surya Karthikeya Mullapudi, Chandravardhan Singh Raghaw, Shahid Shafi Dar, Mohammad Zia Ur Rehman, and Nagendra Kumar. "A multi-temporal multi-spectral attention-augmented deep convolution neural network with contrastive learning for crop yield prediction." *Computers and Electronics in Agriculture* 239 (2025): 110895.

[8]. Nejad, Seyed Mahdi Mirhoseini, Dariush Abbasi-Moghadam, and Alireza Sharifi. "ConvLSTM–ViT: A deep neural network for crop yield prediction using Earth observations and remotely sensed data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024): 17489-17502.

[9]. Mena, Francisco, Deepak Pathak, Hiba Najjar, Cristhian Sanchez, Patrick Helber, Benjamin Bischke, Peter Habelitz et al. "Adaptive fusion of multi-modal remote sensing data for optimal sub-field crop yield prediction." *Remote Sensing of Environment* 318 (2025): 114547.

[10]. Jiang, Hanhui, Liguo Jiang, Leilei He, Bryan Gilbert Murengami, Xudong Jing, Paula A. Misiewicz, Fernando Auat Cheein, and Longsheng Fu. "Yield prediction of root crops in field using remote sensing: A comprehensive review." *Computers and Electronics in Agriculture* 227 (2024): 109600

[11]. Yuan, Jianghao, Yangliang Zhang, Zuojun Zheng, Wei Yao, Wensheng Wang, and Leifeng Guo. "Grain crop yield prediction using machine learning based on UAV remote sensing: A systematic literature review." *Drones* 8, no. 10 (2024): 559.

[12]. Shawon, Sarowar Morshed, Falguny Barua Ema, Asura Khanom Mahi, Fahima Lokman Niha, and H. T. Zubair. "Crop yield prediction using machine learning: An extensive and systematic literature review." *Smart Agricultural Technology* 10 (2025): 100718.

[13]. Sun, Zhenyuan, Boyan Sun, Saiju Li, Zhen Guo, Zhi Yang, Yimin Ding, and Lei Zhu. "Estimation of daily reference crop evapotranspiration in China based on time-space-LSTM model." *Journal of Hydrology* (2025): 133430.

[14]. Hussein, Layth, Prashant Johri, A. Anusha Priya, R. Ramya, J. Karpagam, and A. Devendran. "Crop Yield Forecasting Using Bidirectional Gated Recurrent Unit (Bi-GRU) Networks." In *2025 International Conference on Automation and Computation (AUTOCOM)*, pp. 224-228. IEEE, 2025.

[15]. Saha, Sarmistha, Olga D. Kucher, Aleksandra O. Utkina, and Nazih Y. Rebouh. "Precision agriculture for improving crop yield predictions: a literature review." *Frontiers in Agronomy* 7 (2025): 1566201.

[16]. Jácome Galarza, Luis, Miguel Realpe, Marlon Santiago Viñán-Ludeña, María Fernanda Calderón, and Silvia Jaramillo. "AgriTransformer: A Transformer-Based Model with Attention Mechanisms for Enhanced Multimodal Crop Yield Prediction." *Electronics* 14, no. 12 (2025): 2466.