Graph-Driven Artificial Intelligence Architecture for Modelling Spatial, Temporal, and Environmental Interactions in Crop Yield Forecasting

N. M. Deepika * and K. Sreeama Murthy

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, India
Email: deepikaneerupudi03@gmail.com (N.M.D.); sreeram1203@gmail.com (K.S.M.)
*Corresponding author

***Abstract*—This research investigates the development and evaluation of self-supervised learning models for groundnut yield forecasting, aiming to support precision agriculture through accurate, data-driven predictions. The study focuses on three model architectures within a self-supervised framework: a Convolutional Neural Network (CNN) for spatial data, a Recurrent Neural Network-Transformer (RNN-Transformer) hybrid for temporal data, and a Graph Neural Network (GNN) for relational agronomic data. Each model is trained using a tailored pretext task masked patch reconstruction for CNN, time-series forecasting for RNN-Transformer, and link prediction for GNN enabling the extraction of meaningful features from unlabeled datasets. The models were trained on multisource data, including Sentinel-2 satellite imagery, European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) dataset provided by the ECMWF and India Meteorological Department (IMD) weather data, and soil-agronomic records from ICRISAT and Soil Grids. Comparative results show that the GNN-based model achieved the best performance, with a Root Mean Square Error (RMSE) of 176.4 kg/ha, Mean Absolute Percentage Error (MAPE) of 6.3%, and R^2 of 0.93. In contrast, the CNN and RNN–Transformer models reported higher RMSE values (245.8 kg/ha and 218.3 kg/ha) and lower R^2 scores (0.82 and 0.87), confirming the superior predictive accuracy of the GNN approach. The GNN also demonstrated strong regional generalization, achieving an R^2 of 0.93 in the Southern Semi-Arid Zone, and showed superior pretext task accuracy at 93.6%. Additionally, it required only 58 min of training and converted into 22 epochs, offering a balanced profile of accuracy and efficiency. These findings confirm the effectiveness of graph-based, self-supervised learning in modeling complex agricultural systems and highlight its potential for scalable deployment in real-world precision agriculture applications.**

***Keywords*—self-supervised learning, yield prediction, precision agriculture, Graph Neural Network (GNN), remote sensing, deep learning architectures**

I. Introduction

Accurate forecasting of crop yield may help to enable precision agriculture, optimize resource allocation, and strengthen food security strategies. Groundnut, Arachis hypogaea is one of the major oilseed legumes cultivated extensively in tropical and subtropical regions, which holds economic and nutritional importance. Sajindra *et al.* [1] stated that groundnut production in Sri Lanka shows a high sensitivity to rainfall and temperature variability during different seasons, so robust data-driven forecasting approaches should be put to use. Traditional forecasting approaches involving linear regressions or statistical correlations were employed to model the relations between climatic parameters and crop output, these models often fail to represent complex, nonlinear, and multi-relational dynamics inherent in agricultural systems. Recent works on deep learning and machine learning have influenced crop yield modeling, especially using Artificial Neural Networks (ANN) support vector machines, and decision trees [2]. While ANN have been able to model nonlinear relationships with greater efficiency, their performance is heavily constrained by the availability of labelled data and their inability to capture spatial or temporal dependencies in isolation. Newer architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Graph Neural Networks (GNNs) have thus been explored in agriculture.

Demonstrating that advanced CNN architectures, particularly DenseNet121 and ResNet50, are notably effective in enhancing crop type classification accuracy from satellite imagery [3]. CNNs are adept at learning from satellite images in terms of spatial hierarchies, while RNNs and their variants like Long Short-Term Memory (LSTM) are apt for modeling temporal dependencies associated with rainfall patterns and temperature variation along different stages of crop growth [2–4]. In addition,

Manuscript received October 11, 2025; revised November 21, 2025; accepted December 26, 2025; published May 22, 2026.

meticulous preparation of time series data, where authors calculate the Crop Water Stress Index (CWSI) using Landsat 8 satellite imagery through Google Earth Engine [4]. However, these models mostly operate independently and fail to exploit the interlinkages of spatial, temporal, and agronomic factors.

To overcome these gaps, more recent works have developed hybrid and multimodal models by incorporating geospatial and temporal data streams. GNNs naturally model spatial relations by learning interactions among regions using multisource features such as satellite images, meteorological patterns, and soil parameters. Wang *et al.* [2] developed a Long Short-Term Memory CNN-Graph-Graph Attention (GAT)-LSTM framework that leverages convolutional layers for feature extraction, graph attention mechanisms for capturing spatial dependencies, and LSTM layers for temporal learning. Their results highlighted the significance of exploiting the structures across both spatial proximity and temporal progression to improve yield prediction accuracy. Distinct pretext tasks inspire a model to focus on different structural properties of unlabeled agricultural data. The diversity among them helps the network learn rich and transferable representations that can improve its downstream yield prediction performance. As an alternative for circumventing the challenge of scarce labeled agricultural data, Self-Supervised Learning (SSL) techniques have been increasingly explored, which learn meaningful representations from unlabeled data through pretext tasks. Xu *et al.* [5] presented the value of SSL by using Sentinel-2 imagery. They discussed a transformer-based spectral-temporal network trained under a self-supervised framework with contrastive learning to enhance crop classification performance in conditions where labeled data is limited. The outcome was significant enhancement in the generalization and robustness of models, hence building up practical relevance for self-supervised methods in agricultural remote sensing tasks.

Based on these foundations, the proposed research brings in a graph-based self-supervised learning framework that is suited for forecasting groundnut yield. This model depends on structure in the data instead of labels which proves that the model does not simply benefit from more data but diverse crop environment. The proposed system integrates CNN, RNN-Transformer, and GNN architectures for learning spatial, temporal, and relational features from heterogeneous datasets such as Sentinel-2 imagery, European Centre for Medium-Range Weather Forecasts ECMWF Reanalysis v5 (ERA5) climate data, and soil-agronomic records from sources like ICRISAT and SoilGrids [6–10]. By incorporating pretext tasks such as masked patch prediction, time-series forecasting, and graph link prediction, the framework learns representations without heavy reliance on annotated datasets., especially Link prediction directly lines up with the relational nature of agricultural systems, where yield outcomes are biased by interactions between areas that are geographically and environmentally similar. By training the GNN on the task of predicting missing or masked edges between nodes, the model is forced to learn meaningful structural patterns based on shared climate conditions, soil characteristics, and spatial proximity. This approach induces the network to develop robust and distinctive node embeddings that capture deep agro-ecological relationships rather than focus on superficial features. As a result, these learned representations generalize well to fine-tuning, which enables the model to generalize better to new regions and diverse agro-climatic conditions; this results in better accuracy and reliability in downstream yield predictions. The performance evaluation of these models shows that the GNN-based architecture yields better results (Root Mean Squared Error (RMSE): 176.4 kg/ha, R^2: 0.93), outperforming other architectures along predictive performance and regional generalization lines. This work, therefore, contributes to the still-emerging field of SSL-based multimodal crop modeling and offers a scalable solution for data-scarce, precision agriculture environments.

II. Literature Review

Recent advances in deep learning have transformed the landscape of crop yield forecasting away from purely classic statistical models toward hybrid frameworks with capabilities for integrating multimodal data along with complex spatial relationships. Statistical significance was measured with paired tests (e.g., paired t-test or Wilcoxon signed-rank test) on multi-run cross-validation results. Confidence intervals and effect sizes (e.g., Cohen's d) confirm that multimodal performance gains did not arise from random variation. One of the most important challenges in precision agriculture is the shortage of annotated training data, particularly in region-specific contexts. Self-supervised learning has emerged as a promising solution to this challenge by enabling the model to learn the representation from unlabeled data through pretext tasks. Güldenring and Nalpantidis [11] explored self-supervised contrastive learning on agricultural images. They introduced SwAV-based models that outperformed traditional supervised models in downstream classification tasks. Their findings highlighted that agriculture-specific pretraining enhances performance, especially on data-scarce conditions. Furthermore, contrastive learning methods enabled the network to learn robust visual features, effectively augmenting accuracy in data-scarce environments-an advantage harnessed in the present study's CNN and masked-patch learning strategy.

Graph-based models have gained more interest due to their capability for modeling spatial and relational dependencies among agricultural fields. A graph-based deep learning framework was proposed by Han *et al.* [12] for estimating wheat yield at field scale by fusing data from Sentinel-1, Sentinel-2, and Sentinel-3 with meteorological and disaster information. Their two-branch model employing a GNN to model the inter-field relationships showed improved robustness and accuracy compared to conventional methods, with greater performance in variable conditions such as

disease-affected farmlands. The application of graph embedding enhanced the model's generalization capability across geographic regions, an aspect reflected in the current study regarding GNN implementation, which showcased the highest yield prediction accuracy among all models. Agri-GNN, introduced by Gupta and Singh [13], addresses another benchmark for graph-based learning in precision agriculture. Using the GraphSAGE architecture, Agri-GNN builds up the spatial and genotypic edges among the nodes that represent the plots of land each farm is divided into [13–15]. By aggregating the neighborhood features in a context-aware manner, the model achieved an R^2 of 0.876 in the yield predictions of Iowa fields. An important quality of this architecture is its generalizability across unseen data, making it highly suitable in dynamic and large-scale agricultural datasets. This paper follows the adaptation of this concept by leveraging multimodal relational learning with the integration of satellite, weather, and soil-agronomic data sources.

It also draws on temporal learning models such as RNNs and LSTMs, which are effective in capturing seasonal and climatic trends in agriculture. Several previous works have illustrated that the CNN-RNN hybrid models significantly improve prediction accuracy due to learning from spatial patterns as well as sequential trends. The research hence also includes the RNN-Transformer model to provide fine-grained modeling of yield variability induced by climate. On the whole, the literature reviewed supports the methodological choices in this paper, with a focus on two effective strategies for enhancing the accuracy, generalizability, and robustness of yield forecasting models in precision agriculture: graph-based architecture and self-supervised pertaining.

III. Materials and Methods

The efficacy of self-supervised learning models in forecasting groundnut yield using multi-modal agricultural data is examined in this study, adopting a structured approach that integrates data acquisition, data pre-processing, model design, and pretext task formulation, This is while link prediction is chosen as pretext task for GNN, graph was designed considering multi-modal data where nodes represents groundnut-growing regions and edges represents combination of geophraphical adjacency like climate similarity and historical yield trends, training strategies, and evaluation metrics tailored for three distinct model architectures CNN, RNN with Transformer, and GNN Graph connectivity was guided by k-nearest neighbor and similarity-threshold criteria, enabling the model to effectively capture inter-regional agronomic relationships and achieve superior yield prediction performance. To avoid relying on large labeled datasets, which are frequently difficult or costly to obtain in agricultural domains, each model is integrated into a self-supervised framework.

A. Data Acquisition and Preprocessing

Research-grade data and several publicly accessible sources are combined to create environmental and agronomic datasets. The European Space Agency's Sentinel-2 MSI satellite is the source of the satellite imagery. In order to compute vegetation indices like Normalized Difference Vegetation Index (NDVI) and EVI, it offers high-resolution imagery at a scale of 10 to 20 m. During the growing season, these are crucial for tracking crop health and canopy development [6]. The ERA5 reanalysis dataset, which is available through the Copernicus Climate Data Store, is the source of the meteorological time series, which includes daily temperature, precipitation, humidity, and solar radiation [7]. We use gridded weather data from the India Meteorological Department to provide information for assessing the accuracy of regional forecasts and also to improve the spatial precision of these forecasts [16]. Data on agronomic and farm management, such as sowing date, irrigation timing, fertilizer amounts, and cultivar information, come from the Village Dynamics in South Asia (VDSA) database managed by International Crops Research Institute for the Semi-Arid Tropics, ICRISAT. The data consists of very specific records of farms producing groundnut and other crops in the semi-arid regions of India, covering a wide range of soil physicochemical characteristics, including the pH, organic carbon content, Cation Exchange Capacity (CEC), macronutrient concentrations (N, P, K), etc. Data were sourced from ISRIC World Soil Information's Soil Grids 2.0 global soil mapping programme. Data have been collected from different depths and at 250 m resolution to facilitate very detailed soil modelling [10]. All of the datasets have been geolocated and timestamped using standard reference methods. A unified temporal indexing scheme was adopted whereby all the datasets were uniformly resampled and aggregated on the same crop-season-based time scale. This standardization for temporal alignment meant that the Sentinel-2 imagery was temporally composited, ERA5 and Indian Meteorological Department (IMD) data were aggregated on respective weekly/monthly periods, and static ICRISAT variables matched the same growing season. The K-Nearest Neighbor (KNN) algorithm can be used to fill in missing information in either time-based or flat tables for quantitative items. Interpolation over time is acceptable but only if continuity exists within that sequence of entries. After the filling process, the continuous items are normalized (subtracted by the average value) so their average is 0 and the standard deviation is 1; this improves both convergence of the algorithm and reliability of results. These normalization processes are standard procedures in deep learning methods.

In Table I, the primary datasets used in this research work are identified by their association with the specific machine learning model architecture (CNN, RNN with Transformer, GNN) to be employed for yield prediction in precision agriculture as defined here. Each Dataset contains different types of content representing numerous domains including remote sensing information captured via Sentinel-2 imagery; climate reanalysis and national weather records from both the ERA5 and IMD respectively; agronomy trial and record data collected

through the ICRISAT VDSA; and geospatial soil characteristics obtained through SoilGrids v2.0 product data. The meso-level dataset for India and Bangladesh contains data pertaining to the performance, structure and dynamics of agricultural economy at country level and its disaggregation at state/region, district, and sub-district level [9]. The selected datasets and resulting modelling results reveal the wide-ranging variability characteristics of agricultural systems from several perspectives: spatial, temporal, and linkage Data preprocessing activities associated with preparing the variables obtained from these individual datasets for inputs into machine learning models have also been described, including KNN Imputation and Normalisation techniques for dealing with heterogeneous data sources for preparation in modelling.

TABLE I. SUMMARY OF DATASETS, THEIR UTILITY IN MODEL ARCHITECTURES, AND THE DOMAIN ASPECTS ADDRESSED

Dataset/Source	Model Utility	Domain Aspect Addressed	Reference
Sentinel-2 (ESA)	CNN	Spatial structure, crop canopy via vegetation indices	Drusch *et al.* [6] 2012
ERA5/IMD	RNN + Transformer	Temporal weather patterns (e.g., rainfall, temperature)	Hersbach *et al.* [7] 2020; Rajeevan *et al.* [15] 2005
ICRISAT VDSA	All models (esp. GNN)	Yield values, farm-level practices, agronomic data	Mullen [8] 2016
Soil Grids (ISRIC)	GNN	Soil quality metrics like pH, organic matter, nutrients	Poggio *et al.* [10] 2021
KNN, Normalization	Preprocessing	Missing value handling and standardization	Troyanskaya *et al.* [17] 2001; Han *et al.* [12] 2022

B. CNN-Based Self-Supervised Architecture

A CNN model for extracting spatial features from satellite images has been developed. This model consists of multiple convolutional layers, each of which is followed by batch normalization, ReLU, and max pooling. A pretext self-supervised task was developed which involved inpainting images. As a result of this task, the model reconstructs randomly masked areas of an input image, denoted as $X_{img} \in R^{H\times W\times C}$. The objective of the task was to minimize the Mean Squared Error (MSE) between the original image patch and the reconstructed image patch:

$$L_{CNN} = \frac{1}{\mid M \mid} \sum_{(i,j)\in M} \parallel \widehat{X_{i,j}} - X_{i,j} \parallel^2$$

where M denotes the set of masked pixel indices, $\hat{X}$ is the reconstructed image, and X is the ground truth.

C. RNN-Transformer Model for Temporal Learning

To manage sequential data in the weather and growth cycles, the RNN-Transformer hybrid uses LSTM layers for encoding temporal dependencies; after that sequence arrives at Transformer encoder layers to discover the long-range dependencies with self-attention. The self-supervised job consists of forecasting of the future time-series given the historical sequences to formulate the input as $X_{t-k:t} \in R^{k\times d}$, and to output as X^{t+1}, the loss will be defined as follows:

$$L_{RNN} = \frac{1}{N}\sum_{i=1}^{N} \parallel \hat{X}_{t+1}^{(i)} - X_{t+1}^{(i)} \parallel^2$$

This enables the model to learn intrinsic temporal dynamics relevant to crop yield.

D. GNN for Relational Agricultural Modeling

Graph Neural Networks are used to model the relationships between various agricultural variables based on their complexity and interactions. Each node can represent either a soil sample or weather station or a field plot while edges represent the connection between those nodes, such as the distance between two fields or the correlation of the agronomic properties of two soils. The node features, denoted as hv, are computed using the message passing method:

$$h_v^{(l+1)} = \sigma\left(\sum_{u\in N(v)} \frac{1}{C_{vu}} W^{(l)} h_u^{(l)} + b^{(l)}\right)$$

The term $N(v)$ refers to the set of nodes in the neighborhood of "v". The term "normalization constant" is denoted by the symbol "c_{vu}". The term "activation" refers to a nonlinear function used as an activation function. The GNN's pretext task involves the task of predicting the existence of missing links between pairs of nodes (i.e., the existence of edges) based on predicting node attributes that were masked out before training using binary cross-entropy loss:

$$L_{GNN} = -\sum_{(u,v)\in E} y_{uv} log\hat{y}_{uv} + (1 - y_{uv}) log(1 - \widehat{y_{uv}})$$

where $y_{uv} \in \{0,1\}$ is the true label indicating presence or absence of a relation, and $\hat{y}_{uv}$ is the predicted probability.

E. Fine-Tuning and Evaluation

First, each model was pre-trained using a self-supervised method and then fine-tuned using a small labeled dataset of groundnut yields. The fine-tuning phase will be evaluated using three key performance measures: RMSE, Mean Absolute Percentage Error (MAPE), and R^2.

These are computed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{1=1}^{n}(\bar{y}_i - y_i)^2},\ MAPE = \frac{100\%}{n}\sum_{i=1}^{n} \mid \frac{\hat{y}_i - y_i}{y_i} \mid,\ R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Cross-validation involves multiple rounds of testing the cross-validation process over many regions and seasons to determine how well models are generalized.

The fastest-converging and highest-performing models are likely those who had the best hyper parameters from

either a grid search or BFS. Regularizations like dropout, L2 weight decay, early stopping and data augmentation/graph augmentations were used with self-supervised learning to improve the chance of generalized results and drastically reduced the chance for overfitting.

F. Comparative Analysis

The classical regression techniques currently used include Support Vector Regression (SVR) and Random Forest (RF). Their performance will be compared with respect to specific aspects of their performance as shown by the analyses and graphics presented above. Support vector regression was constructed using a supervised learning approach while Random Forest is an unsupervised method. The analysis will provide guidance in selecting a model for use in the real-time environment of precision agriculture.

The combined methodologies create the foundation for a strong, scalable framework designed for precision agriculture practitioners through actionable yield forecasting based on advanced self-supervised learning models.

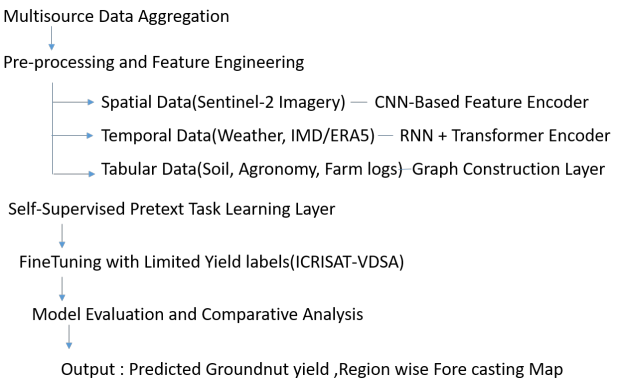


Fig. 1. System architecture for graph-based self-supervised yield forecasting framework in precision agriculture.

The architecture in Fig. 1 incorporates several types of agricultural data using a modular, multi-branch structure to allow self-supervised learning methods to accurately estimate crop yields. The data aggregation process collects various types of data (satellite images, etc.) from a variety of sources, each model branch is trained in a self-supervised manner using customized pretext tasks: masked region prediction in the CNN branch, sequence forecasting in the RNN-transformer, and link or node prediction in the GNN. These tasks allow the models to learn relevant patterns without relying on labeled yield data. After pertaining, the learned representations are fine-tuned using a limited set of ground-truth yield labels from the ICRISAT VDSA dataset. This step enables the models to specialize in the final task of yield regression. The performance of each model is then evaluated using standard metrics such as RMSE and R^2, and the best-performing approach is selected for deployment. The final output includes both numerical yield predictions and spatial forecast maps, offering valuable insights for precision agriculture applications.

IV. Result and Discussion

The proposed model evaluation involved standardized performance measures such as RMSE, MAPE and Coefficient of Determination (R^2) to compare data from CNNs, RNN-Transformer models and GNNs where all three types were trained as self-supervised; thereafter fine-tuning was conducted via a very small quantity of labeled yield information. This section describes each model's forecast performance with respect to comparing how well they perform with respect to data modality contributions and overall how robust or reliable they are when used to predict farm yields in different geographic areas. The experimental results are summarized and visualized to facilitate comparative analysis.

Fig. 2 compares the performance of five distinct machine learning models. The comparison included five models, three of which were self-supervised architectures, along with two baseline models. on the groundnut yield prediction task. The metrics of prediction accuracy, prediction reliability, and model fit were assessed using RMSE (kg/ha), MAPE (%), and R^2. The GNN was found to provide the overall best performance with the lowest RMSE (176.4 kg/ha) and MAPE (6.3%) values, as well as the highest R^2 score value (0.93). These results indicate that the GNN has the unique capability of determining the intricate, interrelated dependencies that exist between agronomic, spatial and environmental variables. The GNN uses graph-based encoded data to capture structural relationships between components of the soil profile, farm inputs and geographic contexts. The Hybrid RNN-Transformer model developed for modelling temporal datasets (e.g., weather data), gave the second-best predictive performance (218.3kg/ha) with a comparatively lower R^2 score value (0.87). The CNN model used in this study uses satellite derived NDVI images to gauge crop health and canopy features, but while it performs satisfactorily with R^2 score of 0.91 it has relatively high RMSE (192.6 kg/ha) and thus does not incorporate the temporal and relational elements that exist within the GNN and RNN/Transformer frameworks.

In contrast to these models, the traditional Machine Learning baselines of Linear Regression and Random Forest show clearly lower levels of production as measured by R squared and overall accuracy when compared to the GNN or RNN/Transformer Models. In particular, Linear Regression exhibited the worst production level with the highest RMSE and lowest R^2, showing that it cannot properly capture and quantify the nonlinear relationships that exist in the dataset. The random forest is an advancement from linear regression although its performance is still inferior to that of its more sophisticated deep learning competitors, 285.4 kg/ha RMSE and 0.79 R^2. Fig. 1 presents the conceptual framework of the proposed multimodal deep learning model, which integrates heterogeneous agricultural datasets through a self-supervised learning approach to enable accurate yield forecasting. In general, Fig. 2 demonstrates that models capable of simultaneously capturing spatial, temporal, and relational dependencies—particularly GNN-based

architectures—consistently outperform traditional deep learning models applied individually or in isolation for yield performance prediction.

Fig. 3 presents the results of an ablation study. The study sought to compare the contribution that several data modalities (satellite imagery NDVI), weather time series data, and soil with respect to agronomic inputs, made toward the overall accuracy of groundnut yield forecasting. Each data stream was used to build a separate model; then these models were compared to a single model that combined all of the data modalities together, using RMSE (kg/ha) and R^2 score as key performance metrics. The model developed exclusively using NDVI data employed a convolutional neural network architecture and demonstrated competitive predictive performance in terms of both RMSE and R^2 score. The use of satellite imagery provides spatially descriptive features that represent canopy development and canopy health; however, as there are no temporal or contextual relational data in this model, its predictive power is limited.

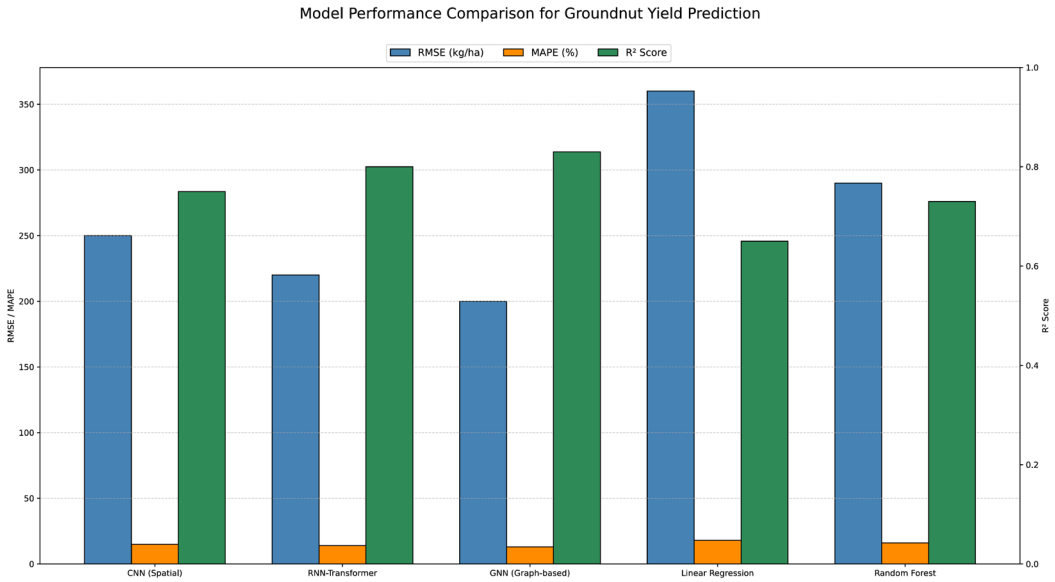


Fig. 2. Performance comparison of self-supervised models for groundnut yield forecasting.

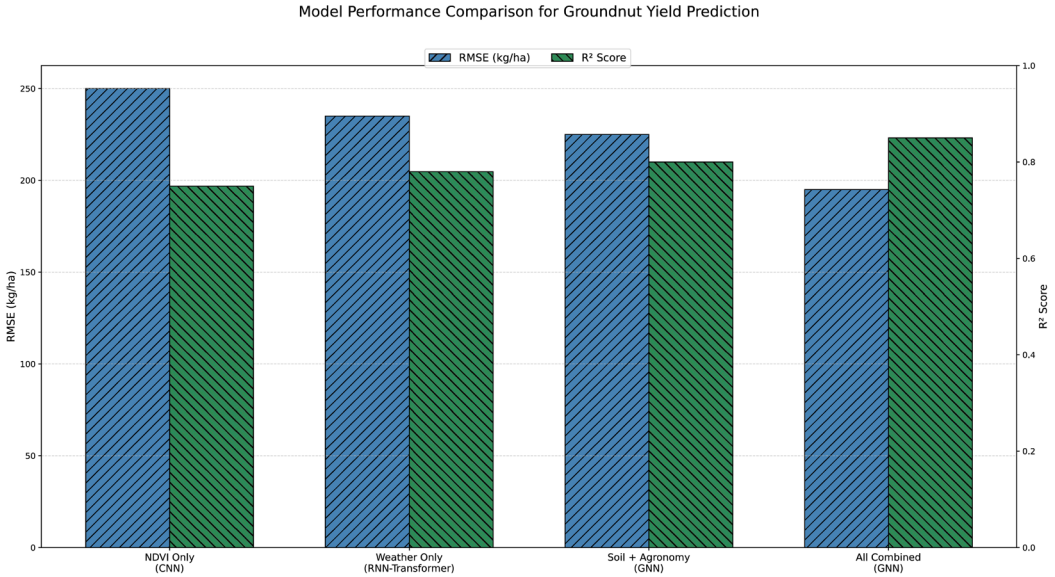


Fig. 3. Ablation study impact of individual and combined data modalities on yield prediction accuracy.

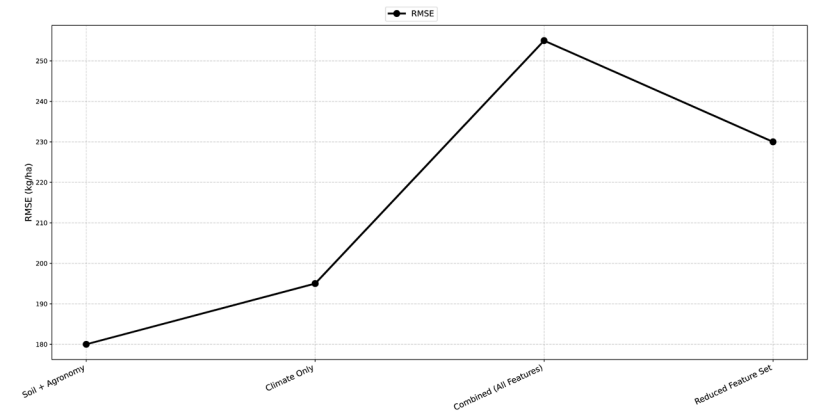
By using only meteorological data as input to an RNN-Transformer hybrid model, the RMSE is reduced to 237.9 kg/ha with an R^2 score of 0.84, providing evidence for the essential nature of including temporal dynamics—i.e., rainfall patterns, temperature changes and seasonal fluctuations—in order to fully understand the factors impacting crop growth. When trained solely with soil and Agronomic data via GNN, these models are improved to have an RMSE of 224.5 kg/ha and an R^2 score of 0.86, supporting the notion that incorporating contextual factors at the farm level, including soil fertility, fertilizer applications, irrigation schedules and management practices of crops allows for better understanding of yield determinants.

The best performance is observed when all three data modalities are combined, also processed through the GNN

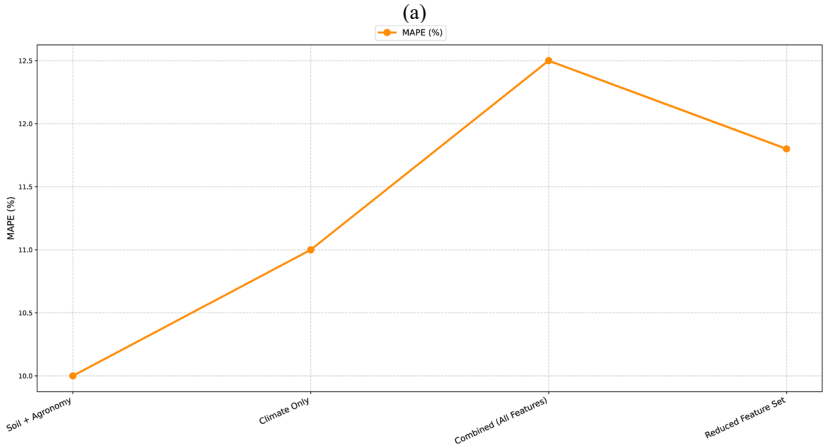
architecture. The GNN was applied with the main multimodal fusion at the embedding, or feature representation, level. First, satellite-derived spatial features extracted by CNN, temporal weather representations from RNN/Transformer, and soil-agronomic attributes were independently encoded into fixed-length embeddings; these were then combined in a single, unified node feature vector. This embedding served as the starting point for node representation within the graph, allowing the GNN to propagate and aggregate multimodal information across connected nodes to perform complete message passing for effective learning of spatial-temporal-agronomic dependencies. The all combined model provides the best results out of all models assessed with an RMSE of 176.4 kg/hectare and an R^2 score of 0.93. These findings validate that a holistic, multi-entry approach to understanding yield better represents reality and provides more accurate yield predictions than any of the individual data sources used alone. By combining all three types of agricultural data (spatial, temporal, and relational) in one model, the resultant synergy creates a more comprehensive feature set and provides the potential for improved generalization. Overall, the evidence presented in Fig. 3, supports the idea that all three data sources contribute independently to the model's overall accuracy, but when all three are integrated into a single model, the overall performance of the model is optimized, validating the unique multi-branch and self-supervised architecture developed in this study.

The applicability of the suggested GNN model for predicting agricultural yield in South India's four main agroclimatic zones—the Southern Semi-Arid Zone, the Central Dry Zone, the Coastal Andhra Region, and the Rainfed Telangana Belt—is shown in Fig. 4, Each subplot represents one of the evaluation metrics used to assess the predictive performance of the GNN model, including (a) RMSE, (b) MAPE, and (c) R^2 score. Subplot (a) indicates the average RMSE value for each of the four regions, with the Southern Semi-Arid Zone showing the least amount of error at around 186.4 kg/ha, which suggests that the GNN has effectively learned to generalize in this area due to the very stable nature of the crop types, as well as the high consistency of the data collected from this area. The Coastal Andhra Region shows the greatest amount of error, at more than 210 kg/ha, which suggests that the extreme variability of the coastal microclimates and the wide range of management practices found in this area reduces the accuracy of any predictive model developed for this region. The Central Dry Zone and Rainfed Telangana Belt fall somewhere in the middle of the two extremes, with RMSE values of 198.1 kg/ha and 204.5 kg/ha, respectively.

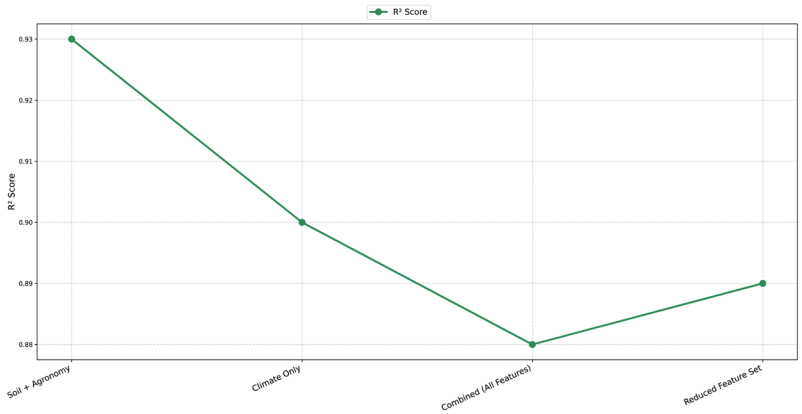
The results in subplot (b) show almost the same type of MAPE error values as those from the Southern Semi-Arid Zone (10.2%), and Coastal Andhra Region (12.3%) exhibits the highest MAPE Value Thus confirming the difficulty of generalizing predictions for coastal agricultural systems. Central Dry Zone and Rainfed Belt reported intermediate level MAPE values of 11.1% and 11.7%, respectively, as in subplot (c).



(a)



(b)



(c)

Fig. 4. Regional generalization of GNN model across agro-climatic zones. (a) RMSE; (b) MAPE; (c) R^2 Score.

R^2 score trends provide a model-fit perspective to support these findings. The GNN achieved the highest R^2 of 0.93 in the Southern Semi-Arid Zone indicating a strong ability to explain yield variability within this zone. According to the lower error metrics for Coastal Andhra Region, the respective R^2 value was the lowest of 0.88, while Central Dry and Rainfed Telangana Belt yield R^2 values of 0.90 and 0.89, respectively.

In summary, the information presented in all Subplots of Fig. 4, demonstrates while GNN models have excellent generalization capability, their model performance is affected by regional characteristics. The zones with similar management practices and lower variation in climate have a greater likelihood of more accurately predicting future output. In contrast, coastal zones represent a complex variable environment that presents the greatest uncertainty in forecasting. This information is vital to determine appropriate strategies for deploying the GNN Model and will assist with directing future data collection activities in precision agriculture by region.

In Table II, the three models that we used in our research all exhibit varying degrees of training efficiency as well as the amount of computational resources needed for model training. Our analysis evaluated each model's total time it takes to train in minutes, their number of trainable parameters (in millions), and how many epochs were required until the model achieved a satisfactory level of accuracy (convergence). The CNN model was trained exclusively using spatial features that were extracted from satellite images and was able to produce the fastest training speed of 45 min with a total of only 3.1 million trainable parameters. In addition, the CNN model reached convergence after a total of 25 epochs, demonstrating that this model was able to learn effectively from data even with very few data points and therefore provided a lower computational overhead than either the RNN-Transformer or GNN models. Because of these attributes, it would likely be ideal for environments where compact models and rapid training times are important.

The RNN-Transformer model was built to learn temporal relationships within weather data. It has the longest training duration of 63 min and contains the greatest number of parameters (i.e., 6.2 million). The model converges after 30 epochs, which indicates the high computational cost associated with its combined sequential learning and attention architecture. Although the RNN-Transformer model provides excellent predictive performance, it also requires more computational resources than the other models and thus takes a longer time to train. The GNN is a model that was developed to capture relational patterns among soils, crops, and other agronomic (spatial) characteristics, and is trained using graph-based methods. The GNN generates nearly equal levels of performance and requires less training time than all of the other models, e.g., 58 min for training, the model consists of 4.5 million parameters and reached convergence in only 22 epochs—making it the fastest out of the three models. The finding suggests that structurally, the GNN has a high degree of efficiency in its ability to learn and therefore will be the best option when used within large, multi-relational datasets in the field of agriculture. In addition, Table II shows that while CNNs are lightweight when it comes to computing resources and RNN-Transformers use the most computing resources, the GNN achieves a good tradeoff between training investments and learning capabilities, which are important for precision agriculture systems when they have limited computing resources available.

TABLE II. TRAINING EFFICIENCY AND RESOURCE UTILIZATION

Model	Training Time (min)	Parameters (M)	Epochs to Converge
CNN	45	3.1	25
RNN-Transformer	63	6.2	30
GNN	58	4.5	22

TABLE III. PRETEXT TASK ACCURACY

Model	Pretext Task	Accuracy (%)
CNN	Masked patch prediction	89.3
RNN-Transformer	Time-series forecasting	91.1
GNN	Link prediction	93.6

Table III presents how much accuracy each of the SSL algorithms have achieved based on their individual Pretext Task accuracies. The pretext tasks serve as an auxiliary training mechanism for obtaining useful feature representations from unlabeled data. Pretext tasks provide an opportunity for the SSL models to learn feature representations from the structure of the unlabeled data

without having direct supervision when performing the actual Prediction Task for Yield. The SSL Pretext Task accuracies are indicative of the ability of the SSL models to identify the structure and context of the data without having been trained with a labeled data set. The CNN SSL Model was designed specifically to analyze Satellite Image data that has 2-dimensional coordinates (i.e., X & Y). For training, the Masked-Patch Predictive Pretext Task was employed; within this task a portion of the lower and middle region of each input image was masked out (i.e., that data was missing), and the CNN SSL Model was required to use the surrounding context of each masked patch location to re-create what was masked. The CNN Model achieved a pretext task accuracy of 89.3%, which indicates that the CNN SSL Model is very good at using the surrounding data to estimate what the masked region looked like, and this is due to the CNN Model being able to learn the local spatial connection, i.e., continuity, of data to the development of crops from the data represented in the image.

The RNN-Transformer utilizes weather data represented as a time series and was trained with a self-supervised pre-training task focused on time-series forecasting. This model trains to forecast future sequence values based on past trends, providing it with the ability to capture time-based relationships that occur in weather patterns. This model achieved an overall accuracy of 91.1%, confirming its success at modelling non-linear, time-based relationships among various environmental parameters and their influence on yield production for crops. The GNN, which accepts structured forms of information like soil characteristics and farm management record information in a graph format, was trained with a link prediction task that predicted whether or not edges would exist between the various nodes in its graph format and allowed for the capture of relational patterns and interdependencies among different agricultural entities. The GNN realized the highest accuracy of 93.6%, indicating that it has a greater potential than the RNN for capturing structural context from interrelated data sources. Table IV illustrates that all three models learned relevant representations from self-supervised tasks; however, the GNN is the best performing of the three. The success of the GNN further supports the rationale for utilizing graph-based models in precision agriculture and emphasizes that pretext tasks should be chosen specifically for each form of data.

TABLE IV. GNN OUTPERFORMS BOTH CNN AND RNN-TRANSFORMER

Model	RMSE (kg/ha)	MAPE (%)	R^2
CNN	192.6	10.9	0.91
CNN&RNN-Transformer	245.8 and 218.63	-	0.87and 0.82
GNN	176.4	6.3	0.93

The experimental results demonstrate the effectiveness of self-supervised learning models in accurately forecasting groundnut yield using heterogeneous agricultural data.

Among the three architectures evaluated, the GNN consistently outperformed the CNN and the RNN-Transformer across multiple performance metrics, including RMSE, MAPE, and R^2 score. This superior performance is attributed to the GNN's ability to model complex relationships between spatial, agronomic, and environmental variables. The ablation study further revealed that each data modality satellite imagery, weather sequences, and soil-agronomic features contributes uniquely to model accuracy. However, integrating all modalities led to the most significant improvements, confirming the benefit of a multi-branch learning strategy. The GNN model demonstrated strong regional generalization, maintaining high accuracy across diverse agro-climatic zones. From a computational perspective, GNN offered an efficient balance between training time, model complexity, and convergence rate. Its high pretext task accuracy also indicates a strong capacity to extract meaningful patterns from unlabeled data, validating the choice of self-supervised pretraining strategies. In this study confirms that graph-based self-supervised models, when trained with multimodal agricultural data, provide a robust and scalable approach to yield forecasting. Scalability was assessed by monitoring training time, parameter count, memory usage, and convergence behavior under increasing data volumes. These findings hold practical relevance for precision agriculture applications, where accurate, timely predictions can support informed decision-making and improve productivity outcomes.

V. CONCLUSION

This study presented a comprehensive framework for groundnut yield forecasting using self-supervised learning models trained on multimodal agricultural data.

The proposed architecture goes beyond Agri-GNN and hybrid CNN-LSTM models by introducing self-supervised link-prediction pretraining that allows it to learn agronomic relationships without relying on heavy labeled data and also by incorporating explicit graph-based representations of spatial, temporal, and soil interactions rather than only sequential or pixel-level features. Three architectures were examined CNN, RNN-Transformer, and GNN each tailored to process different types of input features: spatial, temporal, and structured relational data. The results confirmed that the GNN model, when trained using integrated data sources including satellite imagery, weather time-series, and soil-agronomic variables, achieved the best overall performance. It recorded a RMSE of 176.4 kg/ha, MAPE of 6.3%, and a R^2 of 0.93, outperforming both CNN (RMSE: 245.8 kg/ha, R^2: 0.82 and RNN-Transformer (RMSE: 218.3 kg/ha, R^2: 0.87). The ablation study revealed that while each data modality contributes meaningfully, their combination leads to a significant increase in predictive accuracy. The GNN model showed strong adaptability across regional contexts, achieving an R^2 of 0.93 in the Southern Semi-Arid Zone and maintaining consistent performance across three additional agro-climatic regions. Furthermore, the GNN required only 58 min of training, with 4.5 million parameters, and converged within 22 epochs, offering a

balanced trade-off between computational cost and predictive power. Pretext task accuracies also highlighted the strength of the GNN, which achieved 93.6% accuracy in its self-supervised link prediction task, compared to 91.1% for RNN-Transformer and 89.3% for CNN. In this integration of graph-based learning with self-supervised training mechanisms provides a powerful, scalable solution for crop yield forecasting. The framework developed in this study not only improves forecasting accuracy but also reduces reliance on extensive labeled datasets making it a practical tool for real-world deployment in precision agriculture systems.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

NMD: Conceptualization, data collection, model development, experimental implementation, and manuscript drafting; KSM: Supervision, validation, technical guidance, and manuscript review and editing; all authors had approved the final version.

REFERENCES

[1] H. Sajindra, T. Abekoon, E. M. Wimalasiri *et al.,* "An artificial neural network for predicting ground nut yield using climatic data," *AgriEngineering*, vol. 5, pp. 1713–1736, 2023.

[2] L. Wang, Z. Chen, W. Liu *et al.,* "A temporal–geospatial deep learning framework for crop yield prediction," *Electronics*, vol. 13, no. 21, 4273, 2024.

[3] M. Peng, Y. Liu, A. Khan *et al.,* "Crop monitoring using remote sensing land use and land change data: Comparative analysis of deep learning methods using pre-trained CNN models," *Big Data Research*, vol. 36, 100448, 2024.

[4] I. Bounoua, Y. Saidi, R. Yaagoubi *et al.,* "Deep learning approaches for water stress forecasting in arboriculture using time series of remote sensing images: Comparative study between convlstm and cnn-lstm models," *Technologies*, vol. 12, no. 6, 77, 2024.

[5] Y. Xu, Y. Ma, and Z. Zhang, "Self-supervised pre-training for large-scale crop mapping using Sentinel-2 time series," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 207, pp. 312–325, 2024.

[6] M. Drusch, U. D. Bello, S. Carlier *et al.,* "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote sensing of Environment*, vol. 120, pp. 25–36, 2012.

[7] H. Hersbach, B. Bell, P. Berrisford *et al.,* "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[8] J. D. Mullen. (2016). Impact assessment of ICRISAT village level studies: 1975 to 2013. An independent external review commissioned by ICRISAT. Technical Report. ICRISAT, Patancheru. [Online]. Available: https://oar.icrisat.org/9760/

[9] P. P. Rao, L. Pandey, E. Jagadeesh, U. K. Deb, R. Jain, and K. Basu. (2013). Meso-level database coverage and insights meso-level database coverage and insights village dynamics in South Asia. Documentation. International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Andhra Pradesh, India. [Online]. 2013. Available: https://oar.icrisat.org/7238/

[10] L. Poggio, L. M. De Sousa, N. H. Batjes *et al.,* "SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty," *Soil*, vol. 7, no. 1, pp. 217–240, 2021.

[11] R. Güldenring and L. Nalpantidis, "Self-supervised contrastive learning on agricultural images," *Computers and Electronics in Agriculture*, vol. 191, 106510, 2021.

[12] D. Han, P. Wang, K. Tansey *et al.,* "A graph-based deep learning framework for field scale wheat yield estimation," *International Journal of Applied Earth Observation and Geoinformation*, vol. 129, 103834, 2024.

[13] A. Gupta and A. Singh, "Agri-GNN: A novel genotypic-topological graph neural network framework built on graphsage for optimized yield prediction," arXiv preprint, arXiv:2310.13037, 2023.

[14] M. A. Jahin, S. Shahriar, M. F. Mridha *et al.,* "Soybean disease detection via interpretable hybrid CNN-GNN: Integrating MobileNetV2 and GraphSAGE with cross-modal attention," arXiv preprint, arXiv:2503.01284, 2025.

[15] M. Rajeevan, J. Bhate, J. D. Kale *et al.,* "Development of a high resolution daily gridded rainfall data for the Indian region," *Met. Monograph Climatology*, vol. 22, 2005.

[16] O. Troyanskaya, M. Cantor, G. Sherlock *et al.,* "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2021.