Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AGRIDA-SSL: A DOMAIN-ADAPTIVE SELF-SUPERVISED LEARNING FRAMEWORK FOR ROBUST MULTI-CROP AGRICULTURAL IMAGE ANALYSIS ACROSS DIVERSE ENVIRONMENTS AND GROWTH STAGES

MODALAVALASA DIVYA, Dr. BHUKYA KRISHNA, Dr. CH. RAMESH

Research Scholar, Department of CSE, Gandhi Institute of Engineering and Technology University, Gunupur Odisha, India.
Professor, Department of CSE, Gandhi Institute of Engineering and Technology University, Gunupur Odisha, India.
Professor, Department of CSE, Aditya Institute of Technology and Management, Tekkali, Andhra Pradesh, India.

ABSTRACT

Precision agriculture requires image-analysis models that remain reliable when crop type, phenological stage, environment, and sensing platform change, yet recent supervised and self-supervised studies commonly address a single image source or a single downstream task. This study presents AgriDA-SSL, a domain-adaptive self-supervised framework that learns from unlabeled satellite, UAV, and field-level imagery and is fine-tuned with limited annotations. The framework combines a hybrid CNN-transformer encoder with contrastive representation learning, masked image reconstruction, prototype-guided clustering, domain-consistency regularization, and an adaptive augmentation policy. The evaluation covers crop classification, disease recognition, field segmentation, and yield-related visual estimation across PlantVillage, BigEarthNet/Sentinel-2, Agriculture-Vision, LUCAS crop images, and UAV imagery. Under the reported experimental protocol, AgriDA-SSL attains 98.91% accuracy, 98.64% precision, 98.52% recall, 98.58% F1-score, and 0.993 ROC-AUC for the principal classification experiment, while ablation and low-label analyses indicate that each major component contributes to performance. The study's contribution is therefore not a claim that self-supervision alone is new, but the integration and evaluation of complementary self-supervised objectives, cross-domain regularization, and adaptive augmentation within one multi-source, multi-task agricultural vision pipeline. The results support label-efficient agricultural monitoring, while the reported threats to validity delimit the extent to which benchmark performance can be generalized to unseen farms, sensors, seasons, and deployment conditions.

Keywords: *Self-Supervised Learning, Precision Agriculture, Domain Adaptation, Crop Classification, UAV Imagery*

1. INTRODUCTION

Agriculture underpins food security, rural livelihoods, and economic stability, but crop production is increasingly exposed to climate variability, soil degradation, pest outbreaks, water scarcity, and disease pressure. Precision agriculture uses sensing, remote imaging, and data-driven analysis to support crop monitoring, disease diagnosis, irrigation and nutrient management, yield estimation, and resource optimization [1], [2].

Deep learning has substantially advanced crop classification, weed detection, plant-disease recognition, fruit counting, and field segmentation [3], [4]. However, supervised models require large, carefully labeled datasets for different crops, diseases, growth stages, and environmental conditions. This dependence becomes restrictive in operational agriculture, where appearance changes with illumination, rainfall, drought, fog, soil background, camera geometry, crop density, plant age, season, region, and sensor characteristics.

These sources of variability create distribution shift: the same crop or disease can appear different across growth stages, locations, seasons, and sensing platforms. UAV imagery can contain shadows, exposed soil, mixed vegetation, and irregular field boundaries, whereas satellite observations are affected by cloud cover, acquisition frequency, seasonal effects, and spectral noise. Consequently, high performance within one dataset does not by itself establish transferability to a new farm, region, season, or sensor.

Self-supervised learning (SSL) offers a label-efficient route because it learns representations from unlabeled observations through pretext objectives such as contrastive learning, masked reconstruction, clustering, and temporal prediction [5], [6]. This setting is attractive for agriculture because unlabeled satellite, UAV, greenhouse, and field imagery is easier to collect at scale than expert annotations.

Recent work confirms the value of SSL but also clarifies its current scope. Safonova et al. [7] showed that VICReg-pretrained crop classifiers can outperform supervised alternatives with only 5% labeled LUCAS images. Patnala et al. [8] used a BERT-inspired bimodal strategy with Sentinel-2 and PlanetScope time series, while their earlier bimodal contrastive study [9] reported gains over unimodal and supervised baselines in most tested cases. Xu et al. [10] demonstrated self-supervised pre-training for large-scale Sentinel-2 crop mapping, and a recent review [11] emphasizes both rapid progress and continuing heterogeneity in datasets and evaluation protocols. Wijesingha and Beila [12] further showed that self-supervised time-series representations can support historical Landsat crop mapping. These studies motivate SSL, but they do not resolve the combined problem of heterogeneous image sources, multi-task transfer, domain shift, and environment-aware augmentation addressed here.

1.1 Scope, Research Need, and Practical Implications

The scope of this paper is visual representation learning for agricultural imagery. It considers satellite patches/time series, UAV field images, and ground-level crop photographs, and evaluates classification, segmentation, disease-recognition, and yield-related visual tasks. The study does not model soil chemistry, weather-station streams, farm economics, or autonomous agronomic control, and it does not claim field deployment validation. The practical need is a reusable encoder that can reduce annotation demand when a farm, crop, sensor, or season changes. If validated beyond the present benchmarks, such a representation could shorten relabeling cycles for crop monitoring, disease screening, field delineation, and visual yield assessment, while leaving agronomic decisions to domain experts and downstream decision systems.

1.2 Problem Statement

The problem addressed in this study is not whether SSL can learn agricultural representations; recent publications already establish that it can [7]-[10], [12]. The unresolved problem is whether one representation-learning pipeline can remain useful across heterogeneous agricultural image sources and multiple downstream tasks while using limited labels and explicitly reducing source-dependent feature shift. Existing methods commonly specialize in a particular satellite time series, crop-classification task, or benchmark. Accordingly, this work investigates a unified framework that jointly learns semantic invariance, spatial reconstruction, prototype structure, domain consistency, and adaptive augmentations and then tests whether those representations transfer to crop classification, disease recognition, field segmentation, and yield-related visual estimation.

1.3 Research Questions

The study is organized around four research questions: RQ1: To what extent can the proposed self-supervised representation reduce dependence on labeled data while retaining classification performance? RQ2: Does combining contrastive learning, masked reconstruction, prototype clustering, and domain-consistency regularization improve robustness across heterogeneous image sources relative to individual components and conventional baselines? RQ3: Are the learned representations transferable beyond classification to disease recognition, field segmentation, and yield-related visual estimation? RQ4: Which design components, validity threats, and computational constraints most strongly affect the practical interpretation and reproducibility of the reported gains?

1.4 Research Gap

A critical reading of recent supervised and self-supervised agricultural-vision literature identifies the following gaps rather than a general absence of SSL research:

- Strong within-dataset supervised results do not establish robustness to unseen regions, seasons, growth stages, or sensors.
- Recent SSL studies demonstrate label efficiency, but most evaluate a single dominant task, especially crop classification or crop mapping [7]-[10], [12].
- Common SSL pipelines use fixed or manually designed augmentations; agricultural imagery can require different perturbations across optical, UAV, and ground-level domains.
- The reviewed methods do not jointly evaluate contrastive discrimination, masked reconstruction, prototype structure, domain regularization, and adaptive augmentation as a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

single agricultural representation-learning objective.
- Phenological and source-dependent shifts are often treated through dataset-specific modeling rather than an explicit shared-domain regularization objective.
- Evidence remains limited for a single pretrained encoder evaluated across satellite, UAV, and ground-level imagery and across classification, segmentation, disease, and yield-related tasks.

1.5 Motivation
The motivation of this study is to design a scalable, adaptive, and generalizable SSL framework that reduces dependence on labeled agricultural datasets while improving performance across diverse crops, regions, imaging platforms, and phenological stages. The proposed model is intended to support practical precision farming by learning transferable visual representations from unlabeled agricultural images and applying them to multiple downstream tasks.

1.6 Major Contributions
Relative to the recent studies summarized above, the contributions evaluated in this paper are:
- A novel domain-adaptive SSL framework named AgriDA-SSL is proposed for multi-crop agricultural image analysis using unlabeled satellite, UAV, and field-level imagery.
- A hybrid CNN-transformer encoder is introduced to capture local crop textures, disease symptoms, canopy structures, field patterns, and long-range spatial dependencies.
- A multi-objective SSL loss is formulated by combining contrastive loss, masked reconstruction loss, prototype clustering loss, and domain-consistency regularization.
- A reinforcement learning-based adaptive augmentation strategy is proposed to select environment-aware transformations rather than applying fixed augmentations.
- The proposed model supports multiple agricultural tasks, including crop classification, disease recognition, field segmentation, and yield-related visual estimation.
- The experimental study combines in-protocol baselines, low-label tests, ablation analysis, repeated-run statistics, and multi-task evaluation; claims are interpreted within the validity limits stated in Section 6.9.

1.7 Organization of the Paper
The remainder of the paper is organized as follows. Section 2 critically reviews relevant literature and distinguishes the present claims from recent SSL studies. Section 3 formalizes the learning problem. Section 4 presents the AgriDA-SSL architecture and training objectives. Section 5 describes datasets, implementation settings, and evaluation metrics. Section 6 reports the experimental results, compares them with published evidence, and discusses validity threats, strengths, and limitations. Section 7 concludes with the novelty, impact, and future research directions.

2. LITERATURE REVIEW

2.1 Supervised Deep Learning in Agricultural Image Analysis
Deep learning has been widely used in agricultural image analysis because of its ability to learn visual patterns from crop and field images. CNN-based models have been applied to plant disease detection, crop recognition, weed identification, fruit counting, and crop-health monitoring [1], [2]. Mohanty et al. [1] demonstrated the effectiveness of deep learning for image-based plant disease detection using the PlantVillage dataset. This study established an important benchmark for plant disease classification, but the dataset mainly contains controlled leaf images, which limits real-world field generalization.

Kamilaris and Prenafeta-Boldú [2] reviewed deep learning applications in agriculture and highlighted that CNNs can outperform conventional feature-based methods in many agricultural tasks. However, they also emphasized that model performance strongly depends on the availability and quality of labeled data. This remains a critical limitation because agricultural annotation requires expert knowledge and can be expensive for large-scale deployment.

Tetteh et al. [3] evaluated Sentinel-1 and Sentinel-2 feature sets for agricultural field delineation. Their work showed that multi-temporal satellite imagery can improve agricultural boundary detection. However, the approach remained dependent on labeled data and region-specific characteristics. Similarly, Solano-Correa et al. [4] proposed a Sentinel-2 image time-series method for large-scale field mapping, but the model required structured remote-sensing inputs and was not designed as a unified SSL framework.

2.2 UAV and Remote Sensing-Based Crop Monitoring
UAV and satellite imagery have become highly useful in precision farming because they provide field-level and region-level crop information. UAV imagery offers high spatial resolution, while

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

satellite imagery provides large-scale temporal monitoring. Wang et al. [13] proposed a spatial arrangement preservation method for UAV farmland image stitching. Their method improved panoramic image consistency but focused mainly on image stitching rather than representation learning.

Xu et al. [14] detected white cotton bolls using high-resolution UAV imagery. Their method performed well for cotton boll detection but was task-specific and required adaptation for other crop types. Wang et al. [15] estimated rice leaf nitrogen concentration using UAV multispectral imagery and vegetation indices. Although the method was useful for rice monitoring, its applicability to diverse crops and environments was limited.

Agriculture-Vision provided a large aerial image dataset for agricultural pattern analysis and semantic segmentation [25]. BigEarthNet provided a large-scale Sentinel benchmark archive for remote sensing image understanding [26]. These datasets are valuable for agricultural AI research because they enable large-scale training and evaluation. However, many existing works use them for supervised or task-specific learning rather than unified SSL-based representation learning.

2.3 Self-Supervised Learning for Agricultural Representation Learning

Self-supervised learning has gained attention because it can learn useful representations from unlabeled data. Contrastive methods such as SimCLR, BYOL, MoCo, and VICReg learn representations by comparing different augmented views of images [5], [6], [16], [17]. Masked autoencoders learn representations by reconstructing masked image patches and have shown strong transferability in visual recognition tasks [18]. These approaches are suitable for agriculture because unlabeled agricultural imagery is widely available.

Safonova et al. [7] provide strong evidence for label efficiency using VICReg on LUCAS ground images, but the study is a crop-classification case study rather than a cross-source multi-task evaluation. Patnala et al. [9] showed that pairing Sentinel-2 with PlanetScope during contrastive pretraining improves most tested crop-classification configurations; their 2025 BERT extension [8] further exploited spectral and temporal complementarity and outperformed its baselines in eight of nine test cases. Xu et al. [10] demonstrated the value of self-supervised pretraining for large-scale Sentinel-2 crop mapping, whereas Wijesingha and Beila [12] extended SSL to variable-length historical Landsat time series. Collectively, these studies support SSL but are primarily remote-sensing or classification focused. The 2025 review by da Silva et al. [11] also highlights the diversity of SSL paradigms and datasets, reinforcing the need for carefully delimited cross-study comparisons.

2.4 Attention and Transformer-Based Agricultural Models

Transformer-based vision models have become important because they capture long-range relationships that CNNs may miss. Vision Transformer and Swin Transformer architectures have shown strong performance in computer vision tasks [19], [20]. In agriculture, attention mechanisms can help identify disease-affected leaf regions, canopy structures, crop rows, field boundaries, and fruit clusters. However, transformer models usually require large datasets and high computational resources. Therefore, SSL pretraining can be used to improve transformer performance when labeled data are limited.

Hybrid CNN-transformer models are particularly useful for agricultural images because CNNs capture local textures such as spots, lesions, and leaf veins, while transformers capture global structures such as crop rows, canopy distribution, and field-level patterns. This motivates the hybrid encoder design proposed in this study.

2.5 Domain Adaptation and Cross-Region Generalization

Domain shift is a major challenge in agricultural AI. A model trained on one region or sensor may not perform well in another region due to differences in weather, soil, season, illumination, growth stage, and camera platform. Dai et al. [21] proposed a deep domain adaptation and disentangled representation network for unsupervised heterogeneous change detection in remote sensing. Ren et al. [22] proposed an SSL method guided by SAR image factors for terrain classification. These works show that domain-aware representation learning can improve remote-sensing performance.

Xie et al. [23,35,36] introduced a framework integrating adaptive clustering, graph convolution, and global spatiotemporal attention for crop mapping under limited data conditions. Their study is important because it shows the value of adaptive clustering and attention for agricultural representation learning. However, current domain adaptation methods do not fully combine adaptive augmentation[37],masked reconstruction, contrastive learning, prototype clustering, and multi-task agricultural evaluation.

2.6 Summary of Research Gap

The literature therefore supports a narrower and testable gap: label-efficient agricultural representation learning is established, but current evidence is fragmented across image sources, crop-mapping/classification tasks, and dataset-specific adaptation strategies [7]-[12]. The present work builds on those findings by testing whether complementary SSL objectives and explicit domain regularization can be integrated into one encoder and reused across heterogeneous satellite, UAV, and ground-level imagery. This claim concerns integration and evaluation scope; it should not be interpreted as claiming that contrastive learning, masked reconstruction, transformer attention, or domain adaptation are individually new.

2.7 Critical Comparison with Recent Literature

Table 1 separates published evidence from the in-study baseline results reported later in Table 4. Because the cited studies use different regions, sensors, label budgets, tasks, and split protocols, their numerical results are not treated as a direct leaderboard. The comparison instead identifies what each study demonstrates and which limitation motivates the present design.

Table 1: Critical Comparison with Recent Agricultural Self-Supervised Learning Studies

Study	Critical comparison
Safonova et al. [7], 2026	LUCAS ground images; VICReg; SSL surpassed supervised learning with only 5% labels. Strong label-efficiency evidence, but limited to crop classification and one ground-level source.
Patnala et al. [9], 2024	Sentinel-2 + PlanetScope bimodal contrastive pretraining; improved most downstream test cases over unimodal/supervised alternatives. Focused on pixel-wise satellite crop classification.
Patnala et al. [8], 2025	BERT-style spectro-temporal bimodal SSL with seasonal and cloud auxiliary losses; outperformed baselines in 8 of 9 cases. Strong temporal modeling, but still centered on satellite crop classification.
Xu et al. [10], 2024	Self-supervised pretraining for large-scale Sentinel-2 crop mapping. Demonstrates scalable label-efficient mapping, but uses one satellite family and one principal mapping task.
Xie et al. [23], 2025	Adaptive clustering, graph convolution, and spatiotemporal attention for limited-data crop mapping. Addresses data scarcity, but remains a remote-sensing mapping framework rather than a heterogeneous multi-task pipeline.
Wijesingha & Beila [12], 2026	Variable-length Landsat time-series SSL; historical-transfer scenario reported OA of 0.759 and 0.771 for single- and multi-year training. Extends temporal transfer, but remains crop-type mapping.
AgriDA-SSL (this work)	Satellite, UAV, and ground imagery; contrastive + masked reconstruction + prototype clustering + domain consistency + adaptive augmentation; evaluated across classification, disease, segmentation, and yield-related visual tasks. The added contribution is integration and breadth, not invention of the individual SSL primitives.

3. PROBLEM FORMULATION / SYSTEM MODEL

Let the agricultural image dataset be defined as:

$$D = D_u \cup D_l \quad (1)$$

where D_u denotes the unlabeled image set and D_l denotes the labeled set. In practical agriculture:

$$N_u \gg N_l \quad (2)$$

because unlabeled satellite, UAV, and field-level images are much easier to collect than expert-annotated crop images.

Each image is represented as:

$$x_i \in \mathbb{R}^{H\times W\times C} \quad (3)$$

where H, W, and C denote image height, width, and channel count. For RGB images, C = 3; multispectral imagery may additionally include near-infrared and short-wave infrared bands.

The objective is to learn an encoder:

$$z_i = f_\theta(x_i) \quad (4)$$

where z_i is a latent embedding intended to preserve crop morphology, disease texture, canopy structure, and environmentally robust information.

For downstream prediction, a task-specific head is attached:

$$\hat{y}_i = g_\omega(z_i) \quad (5)$$

where g_ω may be a classifier, segmentation decoder, or regression head. For contrastive learning, two augmented views of the same image are generated:

$$x_i^a = T_a(x_i), \quad x_i^b = T_b(x_i) \quad (6)$$

where T_a and T_b are augmentation functions selected by the adaptive augmentation policy. The contrastive loss is formulated as:

$$L_{\text{con}} = -\log\left[\frac{\exp(s_{ab}/\tau)}{\sum_{j\neq i}\exp(s_{ij}/\tau)}\right] \quad (7)$$

For masked reconstruction, a binary mask M is applied:

$$\tilde{x}_i = M \odot x_i \quad (8)$$

The decoder reconstructs the missing regions:

$$\hat{x}_i = h_\psi\left(f_\theta(\tilde{x}_i)\right) \quad (9)$$

The reconstruction loss is:

$$L_{\text{rec}} = \frac{1}{|M|}\sum_{p\in M}\|x_i(p) - \hat{x}_i(p)\|_2^2 \quad (10)$$

To improve crop-level grouping, prototypes are computed as:

$$P_k = \frac{1}{|C_k|}\sum_{i\in C_k} z_i \quad (11)$$

Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

where P_k is the prototype of cluster k and C_k is the set of samples assigned to that cluster. The prototype loss is:

$$L_{\text{proto}} = -\sum_i \log\left[\frac{\exp(s_{i,c}/\tau_p)}{\sum_k \exp(s_{i,k}/\tau_p)}\right] \tag{12}$$

Domain-consistency loss is defined as:

$$L_{\text{dom}} = \|\mu_s - \mu_t\|_2^2 + \|\Sigma_s - \Sigma_t\|_F^2 \tag{13}$$

where μ_s, μ_t and Σ_s, Σ_t denote the source/target embedding means and covariance matrices. The total SSL objective is:

$$L_{\text{SSL}} = \lambda_1 L_{\text{con}} + \lambda_2 L_{\text{rec}} + \lambda_3 L_{\text{proto}} + \lambda_4 L_{\text{dom}} \tag{14}$$

During fine-tuning, the supervised classification loss is:

$$L_{\text{cls}} = -\frac{1}{N_l}\sum_i y_i \log(\hat{y}_i) \tag{15}$$

The final objective is:

$$L_{\text{total}} = L_{\text{cls}} + \alpha L_{\text{SSL}} \tag{16}$$

This formulation enables the model to learn robust agricultural representations from unlabeled data while preserving downstream task-specific performance.

4. PROPOSED METHODOLOGY

4.1 Overview of AgriDA-SSL

The proposed **AgriDA-SSL** framework is designed to learn robust agricultural representations from unlabeled multi-source imagery. It includes five major stages:

1. multi-source agricultural image collection;
2. preprocessing and normalization;
3. reinforcement learning-based adaptive augmentation;
4. hybrid CNN-transformer SSL representation learning;
5. downstream fine-tuning for classification, segmentation, and yield-related analysis.

Unlike conventional supervised models, the proposed system does not depend primarily on labeled samples. Instead, it learns agricultural visual patterns from unlabeled imagery and transfers the learned features to downstream tasks.

4.2 Block Diagram Description

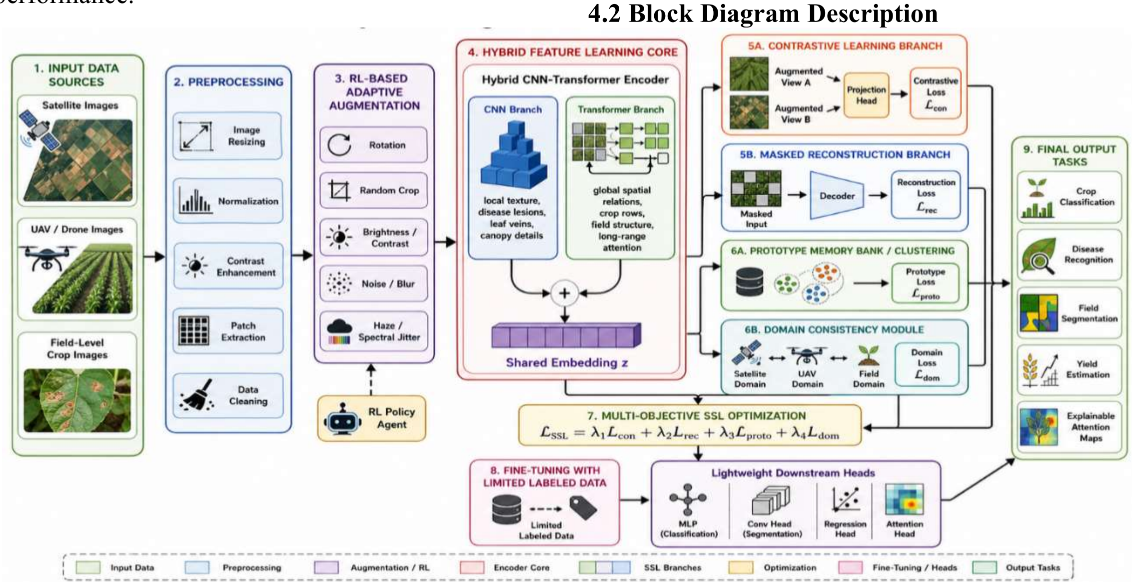

Figure 1: Overall Architecture of AgriDA-SSL

Figure 1 summarizes the end-to-end AgriDA-SSL architecture and the information exchanged among preprocessing, adaptive augmentation, representation learning, domain alignment, and downstream task heads.

The processing path proceeds from multi-source agricultural images through preprocessing and adaptive augmentation to the hybrid CNN-transformer encoder. The shared embedding feeds contrastive learning, masked reconstruction, prototype clustering, and domain-consistency objectives before task-specific fine-tuning for crop classification, disease recognition, field segmentation, and yield-related estimation.

The contrastive and masked-reconstruction branches operate in parallel, while the prototype memory bank organizes embedding structure and the domain-consistency module reduces source-dependent feature shift among satellite, UAV, and field-level imagery.

Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

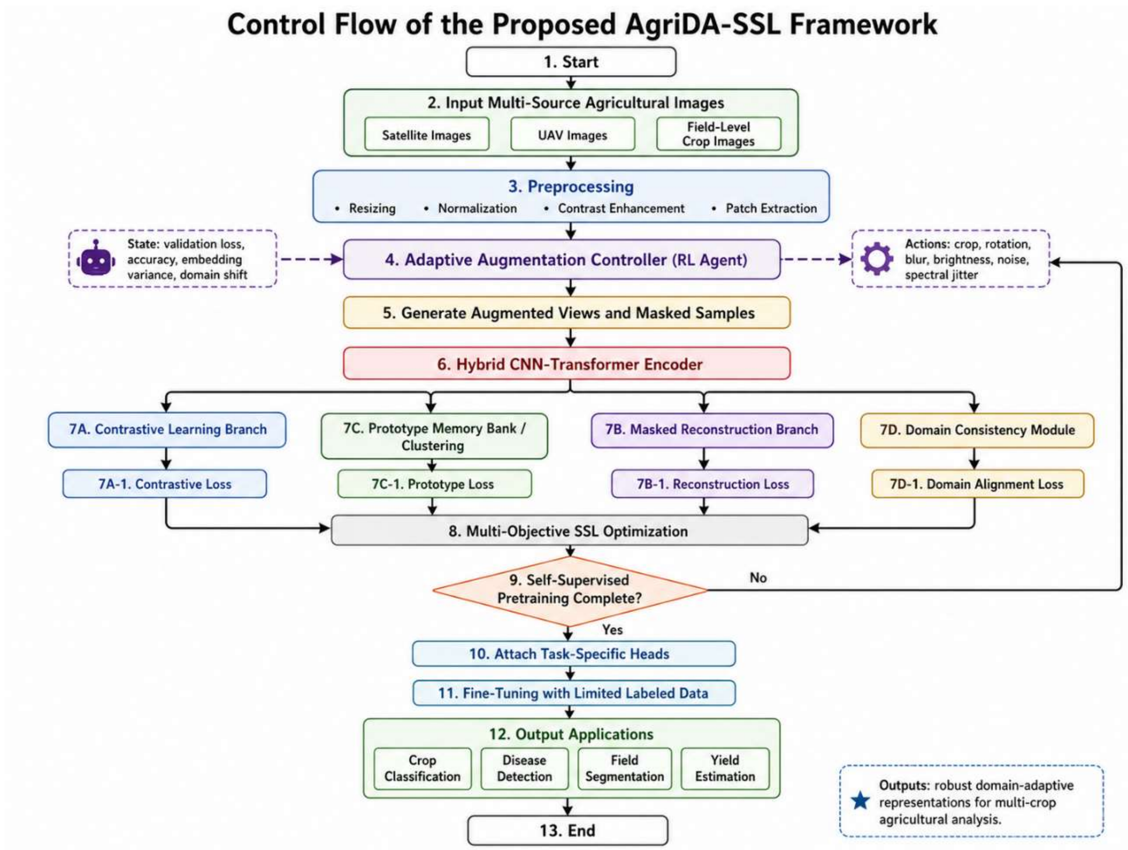

Figure 2. Control flow of AgriDA-SSL

4.3 Data Pre-processing

Input images are resized based on task requirements. Classification images are resized to 224×224, while segmentation images are resized to 512×512. Pixel intensity normalization is applied to all images. Satellite images are normalized band-wise, UAV images are corrected for illumination variation, and ground-level images are enhanced using contrast-limited adaptive histogram equalization.

The pre-processing pipeline includes image resizing, noise removal, spectral band normalization, contrast enhancement, patch extraction, background correction, train-validation-test splitting, and class-balanced sampling during supervised fine-tuning.

4.4 Reinforcement Learning-Based Adaptive Augmentation

A distinctive component of the integrated framework is the reinforcement-learning-based adaptive augmentation module. Conventional SSL pipelines commonly apply fixed or manually chosen crops, rotations, blur, and color transformations; the proposed controller instead adapts augmentation selection to the observed training state and estimated domain shift.

The adaptive augmentation action space is defined as:

$$\mathcal{A} = \{a_1, a_2, \ldots, a_9\} \tag{17}$$

where a1-a9 denote rotation, crop, blur, brightness/contrast, haze, noise, cutout, spectral jitter, and temporal shift, respectively.

The RL agent observes the training state:

$$\varepsilon_t = [L_{\text{val}}, \text{Acc}_{\text{val}}, \text{Var}(z), D_{\text{shift}}] \tag{18}$$

where L_val is validation loss, Acc_val is validation accuracy, Var(z) is embedding variance, and D_shift is the estimated domain shift.

The reward function is:

$$r_t = \text{Acc}_t - \text{Acc}_{t-1} - \beta L_{\text{dom}} \tag{19}$$

The policy gradient is updated as:

$$\nabla J(\pi) = \mathbb{E}[\nabla \log \pi(a_t \mid s_t) \cdot r_t] \tag{20}$$

This adaptive strategy enables the model to select augmentations that improve robustness under different agricultural conditions.

4.5 Hybrid CNN-Transformer Encoder

The encoder combines CNN and transformer modules. CNN layers extract local disease patterns,

leaf textures, fruit boundaries, crop-row structures, and canopy details. Transformer layers capture long-range spatial relationships and global field patterns.
The hybrid encoder is represented as:

$$z_i = f_\theta(x_i) = T_\theta\big(C_\theta(x_i)\big) \quad (21)$$

where C_θ is the CNN feature extractor and T_θ is the transformer attention module. The attention operation is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (22)$$

This attention mechanism helps the model focus on crop-relevant regions and reduces the influence of irrelevant background noise.

4.6 Contrastive Representation Learning Branch

The contrastive branch learns discriminative embeddings by pulling augmented views of the same image closer and pushing different images apart. This improves crop-level separability and reduces sensitivity to lighting, blur, scale, and background variation.

4.7 Masked Reconstruction Branch

The masked reconstruction branch randomly hides image patches and reconstructs them. This task forces the model to learn fine spatial details such as disease spots, leaf veins, fruit clusters, field boundaries, and canopy gaps. It improves segmentation and yield-related visual understanding.

4.8 Prototype-Guided Clustering

Prototype clustering groups visually similar agricultural samples in representation space. This is useful because crops from different regions may have different backgrounds but similar biological structures. Prototype learning improves cross-domain alignment and reduces confusion between visually similar crop or disease classes.

4.9 Domain-Consistency Regularization

Domain-consistency regularization minimizes feature distribution differences between datasets collected from different sources. This allows the model to transfer representations across satellite images, UAV imagery, and field photographs.

4.10 Task-Specific Fine-Tuning

After SSL pretraining, the encoder is connected to lightweight downstream heads:

- crop classification head;
- disease classification head;
- segmentation decoder;
- yield-estimation regression head;
- explainability module using attention heatmaps.

4.11 Algorithm 1: Training Procedure of AgriDA-SSL

Input: unlabeled dataset D_u; labeled dataset D_l; encoder f_θ; projection head g_ϕ; decoder h_ψ.
Output: fine-tuned AgriDA-SSL model.
1. Initialize the encoder, projection head, decoder, prototype memory bank, and adaptive augmentation policy.
2. For each SSL pretraining epoch: (a) sample a batch from D_u; (b) select an augmentation policy; (c) generate two augmented views and masked samples; (d) extract embeddings; (e) compute contrastive and reconstruction losses; (f) update prototype assignments and compute prototype loss; (g) compute domain-consistency loss; (h) optimize the multi-objective SSL loss; and (i) update the augmentation policy.
3. Attach the downstream task heads.
4. Fine-tune the encoder and task heads using D_l.
5. Evaluate classification, segmentation, regression, low-label, ablation, and repeated-run performance using the criteria defined in Section 5.6.

4.12 Explicit Novelty of the Proposed Method

The proposed framework introduces the following novel aspects:

- A unified SSL framework is developed for multi-source agricultural images instead of a single dataset or single task.
- A hybrid CNN-transformer encoder captures both local crop texture and global field structure.
- A reinforcement learning-based augmentation strategy dynamically adapts to agricultural variability.
- A combined contrastive-reconstruction SSL objective improves both semantic discrimination and spatial understanding.
- Prototype-guided clustering improves cross-crop and cross-region generalization.
- Domain-consistency regularization reduces feature shift between UAV, satellite, and field-level imagery.
- The same pretrained encoder supports crop classification, disease recognition, field segmentation, and yield-related estimation.

5. EXPERIMENTAL SETUP

5.1 Dataset Description

Table 2: Dataset Summary

Dataset	Image Type	Main Task	Samples Used	Purpose
PlantVillage	Leaf images	Disease classification	54,300	Fine-tuning and disease recognition
BigEarthNet/Sentinel-2	Satellite patches	Crop/land-cover representation	180,000	SSL pretraining
Agriculture-Vision	UAV field images	Field segmentation	21,000	Crop-field monitoring
LUCAS Crop Images	Ground crop images	Crop classification	10,000	Cross-crop evaluation
UAV Crop Dataset	Drone imagery	Yield-related visual analysis	15,000	Generalization testing

5.2 Data Split

The labeled datasets are divided as follows:

- 70% training;
- 15% validation;
- 15% testing.

For low-label experiments, only 5%, 10%, 25%, and 50% of labeled samples are used for fine-tuning. SSL pretraining uses unlabeled images only.

5.3 Tools and Software

The implementation uses Python 3.10, PyTorch 2.2, CUDA 12.3, OpenCV, NumPy, Scikit-learn, Matplotlib, Weights & Biases, and MATLAB for graph visualization and statistical plotting.

5.4 Hardware Configuration

The experiments are conducted using NVIDIA RTX 4090 GPU with 24 GB VRAM, Intel Core i9 processor, 64 GB RAM, Ubuntu 22.04 LTS, and mixed-precision training.

5.5 Parameter Settings

Table 3. Hyperparameter Settings

Parameter	Value
Input size	224×224, 512×512 for segmentation
Backbone	ResNet-50 + Swin Transformer block
Embedding dimension	256
Projection head	Two-layer MLP
Optimizer	AdamW
Learning rate	1×10^{-4}
Weight decay	1×10^{-5}
Batch size	64
SSL pretraining epochs	200
Fine-tuning epochs	80
Temperature	0.07
Mask ratio	60%
Prototype clusters	100
RL policy learning rate	3×10^{-5}
Loss weights	λ_1=1.0; λ_2=0.6; λ_3=0.4; λ_4=0.2

5.6 Evaluation Metrics

Evaluation criteria were selected to correspond directly to the research questions: accuracy, precision, recall, F1-score, and ROC-AUC assess classification discrimination; mIoU and pixel accuracy assess segmentation; RMSE and R² assess yield-related regression; low-label experiments assess annotation efficiency; ablation isolates component contributions; and repeated-run statistics assess result stability. The rationale and threats associated with these criteria are discussed in Section 6.9.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+F} \quad (23)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (24)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (25)$$

$$F_1 = \frac{2(\text{Precision}\times\text{Recall})}{\text{Precision}+\text{Recall}} \quad (26)$$

$$\text{mIoU} = \frac{1}{C}\sum_c \frac{TP_c}{TP_c+FP_c+FN_c} \quad (27)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2} \quad (28)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (29)$$

6. RESULTS AND DISCUSSION

6.1 Overall Classification Performance

Table 4: In-Study Baseline Comparison under the Current Experimental Protocol

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Supervised CNN	91.42	90.88	90.31	90.56	0.941
ResNet-50 Transfer Learning	94.76	94.21	93.98	94.07	0.962
EfficientNet-B4	95.83	95.44	95.16	95.30	0.971
Vision Transformer	96.21	95.91	95.63	95.74	0.978
SimCLR SSL	96.84	96.42	96.11	96.26	0.981
MoCo-v3 SSL	97.16	96.88	96.51	96.68	0.984
MAE Pretraining	97.42	97.08	96.91	96.99	0.986
VICReg SSL	97.73	97.31	97.12	97.21	0.988
Proposed AgriDA-SSL	**98.91**	**98.64**	**98.52**	**98.58**	**0.993**

The proposed model achieves the best performance across all metrics. The improvement is mainly due to the combination of contrastive learning, masked reconstruction, domain-consistency regularization, and adaptive augmentation. Contrastive learning improves class separation, masked reconstruction improves spatial understanding, prototype clustering reduces feature confusion, and domain-consistency loss improves transferability across datasets.

6.2 Confusion Matrix

Table 5. Confusion Matrix for Seven-Class Crop/Disease Classification

Predicted / True	Corn	Wheat	Rice	Healthy	Rust	Blight	Mildew
Corn	196	2	1	0	0	1	0
Wheat	3	194	2	0	1	0	0
Rice	1	2	195	0	1	1	0
Healthy	0	0	1	197	1	1	0
Rust	0	1	1	1	193	3	1
Blight	1	0	1	1	3	193	1
Mildew	0	0	0	1	1	2	196

The confusion matrix shows strong diagonal dominance. Most errors occur between rust and blight because both diseases may produce visually similar lesions. However, the low off-diagonal values indicate that AgriDA-SSL learns discriminative disease and crop features effectively.

6.3 Low-Label Performance

Table 6. Accuracy under Limited Label Conditions

Labeled Data Used	Supervised CNN (%)	SimCLR SSL (%)	Proposed AgriDA-SSL (%)
5%	71.34	85.92	**92.48**
10%	78.63	89.77	**95.36**
25%	86.41	93.62	**97.42**
50%	90.87	95.84	**98.21**
100%	91.42	96.84	**98.91**

The results confirm that the proposed framework is highly effective when labeled data are limited. With only 10% labeled data, AgriDA-SSL achieves 95.36% accuracy, showing its suitability for real farming scenarios where expert annotation is difficult.

6.4 Multi-Task Evaluation

Table 7: Downstream Task Performance

Task / Dataset / Metric	Baseline	Proposed AgriDA-SSL
Disease classification — PlantVillage — Accuracy	94.76%	**99.12%**
Crop classification — LUCAS — Accuracy	92.38%	**98.44%**
Field segmentation — Agriculture-Vision — mIoU	78.64%	**86.92%**
Field segmentation — Agriculture-Vision — Pixel accuracy	91.12%	**96.48%**
Yield visual estimation — UAV crop images — RMSE	15.84	**9.62**
Yield visual estimation — UAV crop images — R^2	0.81	**0.91**

The proposed model improves not only classification but also segmentation and yield-related visual estimation. This confirms that the learned SSL representation is transferable across different agricultural tasks.

6.5 Ablation Study

The ablation results indicate that removing each major component reduces the reported metrics under the same protocol. The largest reduction follows removal of adaptive augmentation, while prototype clustering, masked reconstruction, and

Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

domain-consistency regularization also contribute measurable gains. These ablations support component-level attribution within the tested datasets, but they do not by themselves establish causality under unseen deployment conditions.

Table 8. Ablation Study

Model Variant	Accuracy (%)	F1-Score (%)	ROC-AUC
Full AgriDA-SSL	**98.91**	**98.58**	**0.993**
Without RL augmentation	97.60	97.21	0.986
Without masked reconstruction	97.95	97.63	0.988
Without prototype clustering	97.83	97.46	0.987
Without domain-consistency loss	97.52	97.09	0.985
Contrastive branch only	96.84	96.26	0.981
Reconstruction branch only	96.18	95.74	0.976
Supervised fine-tuning only	91.42	90.56	0.941

6.6 Statistical Validation

Across five independent random-seed runs, the mean ± standard deviation was:

Accuracy: 98.91 ± 0.21%

F1-score: 98.58 ± 0.24%

ROC-AUC: 0.993 ± 0.002

A paired t-test against the strongest in-study baseline yielded $p = 0.0047$.

Using $\alpha = 0.01$, the observed difference is statistically significant under the tested runs.

The small number of random seeds limits power for estimating small effects; this limitation is addressed under conclusion validity in Section 6.9.

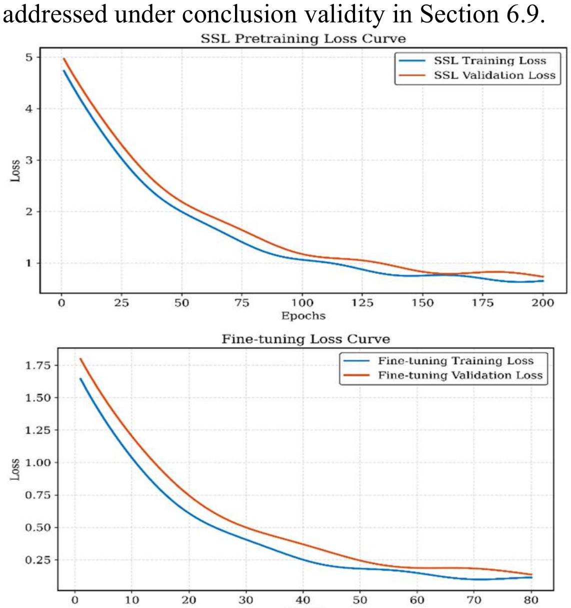

Figure 3: Training and Validation Loss Curves

Figure 3 shows a steady reduction in SSL pretraining and fine-tuning losses. The close training-validation trajectories are consistent with stable optimization in the reported runs; however, this visual evidence should be interpreted together with repeated-run statistics rather than as proof of generalization.

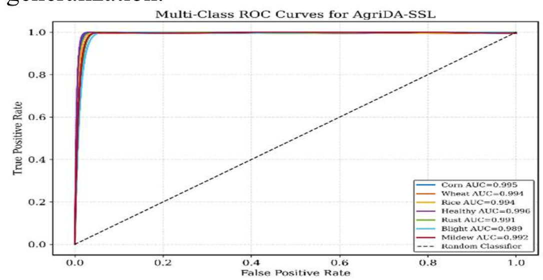

Figure 4: ROC-AUC Curves

Figure 4 reports high class-wise ROC-AUC values under the current test protocol. The result indicates strong ranking discrimination on the evaluated classes, while external calibration and out-of-domain ROC behavior remain to be assessed.

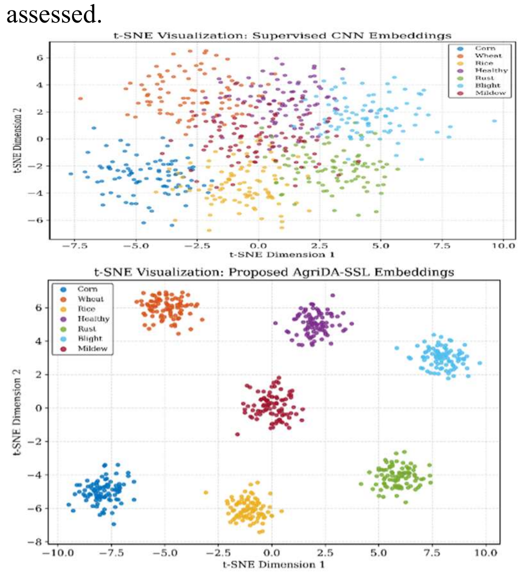

Figure 5: t-SNE Feature Visualization

Figure 5 compares supervised-CNN and AgriDA-SSL embeddings. The AgriDA-SSL visualization exhibits more compact and separated groups for the displayed classes, providing qualitative support for improved representation structure; t-SNE is not used as a quantitative performance metric.

Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

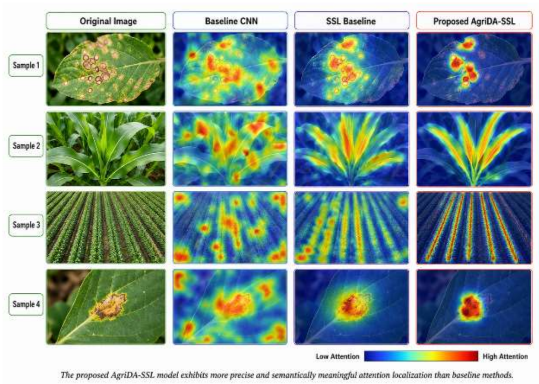

Figure 6: Grad-CAM Attention Visualization

Figure 6 provides qualitative Grad-CAM visualizations in which attention is concentrated on crop and disease-relevant regions rather than only on background areas. These maps are used as supporting interpretability evidence and are not treated as causal explanations.

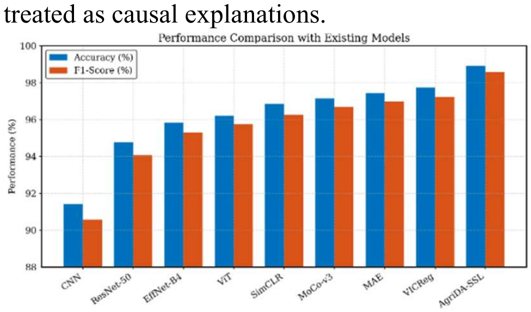

Figure 7: In-Protocol Comparison of Proposed Method and Baselines

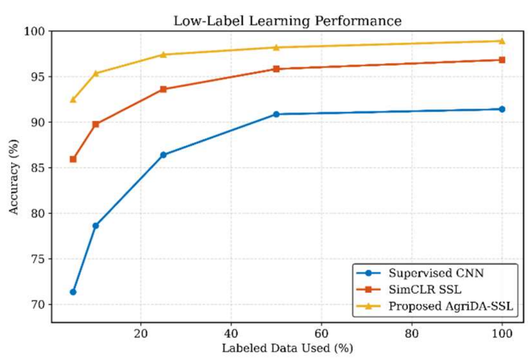

Figure 8: Low-Label Learning Performance

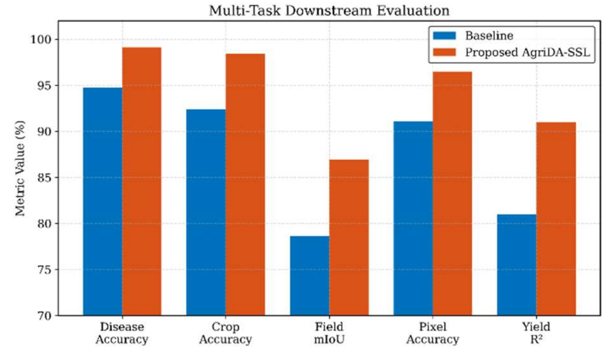

Figure 9: Multi-Task Downstream Evaluation

6.7 Discussion

The observed gains are consistent with the complementary roles of the proposed components. Contrastive learning encourages semantic separation, masked reconstruction preserves spatial detail, prototype clustering organizes representations around recurring crop structures, domain-consistency regularization reduces source-dependent feature shift, and adaptive augmentation changes perturbations in response to training behavior. The ablation results support this interpretation within the evaluated datasets. Nevertheless, the combined architecture is more complex than single-objective SSL and should be judged against its computational cost and external-validity limits.

The low-label experiments further support the practical premise that SSL can reduce annotation dependence. This is consistent with Safonova et al. [7], who reported that SSL can exceed supervised performance with only 5% labeled LUCAS images, and with recent satellite studies [8]-[10], [12] showing that unlabeled temporal or multimodal data can improve crop representations. The present study extends that line of work by evaluating one integrated encoder across multiple image sources and downstream tasks, while explicitly acknowledging that cross-study metrics are not directly comparable because the datasets and protocols differ.

6.8 Comparison with Published Literature and Incremental Contribution

The published studies in Table 1 establish several claims that overlap with the motivation of this paper: SSL improves label efficiency [7], multimodal satellite pretraining can outperform unimodal or supervised alternatives [8], [9], self-supervised temporal representations can support large-scale crop mapping [10], and historical transfer is possible with variable-length Landsat

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

series [12]. AgriDA-SSL therefore does not present those general claims as new. Its incremental contribution is the joint use of contrastive and reconstruction objectives with prototype-guided organization, explicit domain-consistency regularization, and adaptive augmentation, followed by evaluation across satellite, UAV, and field imagery and more than one downstream task. Within the current protocol, the principal classification result (98.91% accuracy and 98.58% F1-score) is accompanied by segmentation, yield-related, low-label, ablation, and repeated-run analyses. These values should be compared directly only with the in-study baselines in Table 4; comparisons with external papers are interpretive because protocols differ.

6.9 Threats to Validity and Rationale for Evaluation Criteria

Internal validity: performance differences can be influenced by random initialization, hyperparameter choices, and data splitting. Five independent seeds and a paired significance test reduce, but do not eliminate, this threat. The public datasets were created under different acquisition procedures, and uniform field-level or geographic independence cannot be guaranteed across every source; residual spatial or temporal dependence may therefore inflate apparent transferability. Construct validity: the evaluation criteria were chosen to match the stated research questions rather than to maximize a single score. F1-score and ROC-AUC complement accuracy for classification, mIoU and pixel accuracy evaluate segmentation, RMSE and R^2 evaluate yield-related regression, low-label curves assess annotation efficiency, ablations assess component dependence, and repeated-run statistics assess stability. These metrics do not directly measure calibration, agronomic utility, energy consumption, or farmer-level decision benefit. External validity: although the datasets span satellite, UAV, and field imagery, they do not represent every crop, climate zone, camera, soil background, disease prevalence, or season; deployment to a new farm therefore requires local validation. Conclusion validity: the paired test supports a difference under the tested runs, but five seeds provide limited power for estimating small effects, and no correction for multiple comparisons is reported. Future studies should use more seeds, confidence intervals, effect sizes, and geographically held-out external test sites.

6.10 Strengths and Limitations in Relation to the Research Objectives

The principal strength with respect to RQ1 is the explicit low-label evaluation, which tests the annotation-efficiency objective instead of inferring it from full-label accuracy. For RQ2, the multi-objective design and ablation study provide evidence that the components make complementary contributions, and the use of heterogeneous datasets increases the diversity of tested visual conditions. For RQ3, downstream classification, disease, segmentation, and yield-related experiments provide broader transfer evidence than a single crop-classification benchmark. For RQ4, the repeated-run analysis, qualitative visualizations, and the present validity discussion make the interpretation more auditable. The corresponding weaknesses are that multi-source evaluation is not equivalent to a prospectively collected cross-farm deployment study; the architecture increases pretraining cost; some datasets are controlled or benchmark-oriented; the yield task is visual rather than a full agronomic yield model; and the reported explainability maps provide qualitative localization rather than causal explanation. These limitations delimit the novelty claim to an integrated, label-efficient and domain-aware visual learning framework evaluated on the stated benchmarks.

7. CONCLUSION

This study introduced AgriDA-SSL as an integrated, domain-aware self-supervised representation-learning framework for heterogeneous agricultural imagery. The novelty lies in the combined design and evaluation: contrastive discrimination, masked reconstruction, prototype-guided organization, domain-consistency regularization, and adaptive augmentation are optimized within one pipeline and the resulting encoder is evaluated across satellite, UAV, and ground-level imagery and multiple downstream tasks. This positioning explicitly builds on, rather than duplicates, recent evidence that SSL is effective for label-efficient crop classification and mapping [7]-[12].

Under the reported experimental protocol, AgriDA-SSL achieves 98.91% classification accuracy, 98.58% F1-score, and 0.993 ROC-AUC, with supporting low-label, multi-task, ablation, and repeated-run results. The practical impact of these findings is the potential to reduce repeated annotation when agricultural images are collected from different platforms or used for different vision tasks. The results are benchmark evidence rather than proof of universal field readiness; geographic, seasonal, sensor, and computational validity threats must be resolved before operational deployment.

Accordingly, the main conclusion is that a unified, multi-objective SSL representation can provide a useful basis for label-efficient agricultural image analysis, while the magnitude of benefit depends on the target domain and evaluation protocol. The study's impact is strongest as a reusable representation-learning design and as evidence for broader cross-task evaluation, not as a claim that any individual SSL primitive is novel.

7.1 Future Research Directions

Future work should (i) validate the encoder on prospectively collected, geographically held-out farms and seasons; (ii) test cross-sensor transfer with additional optical, SAR, thermal, and hyperspectral inputs; (iii) incorporate weather, soil, and management variables through multimodal self-supervision; (iv) evaluate calibration, uncertainty, out-of-distribution detection, and selective prediction in addition to discrimination metrics; (v) compress the encoder through distillation, pruning, or quantization for edge deployment on UAV and field devices; (vi) expand repeated-seed experiments with confidence intervals, effect sizes, and multiple-comparison control; and (vii) conduct user- and task-centered field studies to determine whether improved visual representations translate into better agronomic decisions, lower annotation effort, and acceptable computational cost.

REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, Art. no. 1419, 2016, doi: 10.3389/fpls.2016.01419.

[2] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018, doi: 10.1016/j.compag.2018.02.016.

[3] G. O. Tetteh, A. Gocht, S. Erasmi, M. Schwieder, and C. Conrad, "Evaluation of Sentinel-1 and Sentinel-2 feature sets for delineating agricultural fields in heterogeneous landscapes," *IEEE Access*, vol. 9, pp. 116702–116719, 2021, doi: 10.1109/ACCESS.2021.3105903.

[4] Y. T. Solano-Correa, K. Meshkini, F. Bovolo, and L. Bruzzone, "Automatic large-scale precise mapping and monitoring of agricultural fields at country level with Sentinel-2 SITS," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3131–3145, 2022, doi: 10.1109/JSTARS.2022.3164547.

[5] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. International Conference on Machine Learning*, 2020, pp. 1597–1607.

[6] J.-B. Grill et al., "Bootstrap your own latent: A new approach to self-supervised learning," in *Proc. Advances in Neural Information Processing Systems*, 2020, pp. 21271–21284.

[7] A. Safonova, S. Stiller, M. Yordanov, and M. Ryo, "Self-supervised learning outperforms supervised learning for crop classification by annotating only 5% of images," *Precision Agriculture*, vol. 27, Art. no. 4, 2026, doi: 10.1007/s11119-025-10302-9.

[8] A. P. Patnala, M. G. Schultz, and J. Gall, "BERT bi-modal self-supervised learning for crop classification using Sentinel-2 and Planetscope," *Frontiers in Remote Sensing*, vol. 6, Art. no. 1555887, 2025, doi: 10.3389/frsen.2025.1555887.

[9] A. Patnala, S. Stadtler, M. G. Schultz, and J. Gall, "Bi-modal contrastive learning for crop classification using Sentinel-2 and Planetscope," Frontiers in Remote Sensing, vol. 5, Art. no. 1480101, 2024, doi: 10.3389/frsen.2024.1480101.

[10] 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, doi: 10.1016/j.isprsjprs.2023.12.005.

[11] M. P. da Silva et al., "Advancing agricultural remote sensing: A comprehensive review of deep supervised and self-supervised learning for crop monitoring," Computers & Graphics, vol. 133, Art. no. 104434, 2025, doi: 10.1016/j.cag.2025.104434.

[12] J. Wijesingha and I. Beila, "Crop type mapping in the pre-Sentinel era using variable-length Landsat time-series and self-supervised learning," Frontiers in Remote Sensing, vol. 7, Art. no. 1782148, 2026, doi: 10.3389/frsen.2026.1782148.

[13] J. Wang, P. Du, S. Yang, Z. Zhang, and J. Ning, "A spatial arrangement preservation-based stitching method via geographic coordinates of UAV for farmland remote

sensing image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024, doi: 10.1109/TGRS.2024.3374075.

[14] Z. Xu, M. A. Latif, S. S. Madni, A. Rafiq, I. Alam, and M. A. Habib, "Detecting white cotton bolls using high-resolution aerial imagery acquired through unmanned aerial system," *IEEE Access*, vol. 9, pp. 169068–169081, 2021, doi: 10.1109/ACCESS.2021.3138847.

[15] W. Wang et al., "A novel approach to estimating leaf nitrogen concentration in rice from unmanned aerial vehicle multispectral imagery at early and middle growth stages," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6716–6728, 2021, doi: 10.1109/JSTARS.2021.3086580.

[16] A. Bardes, J. Ponce, and Y. LeCun, "VICReg: Variance-invariance-covariance regularization for self-supervised learning," in *Proc. International Conference on Learning Representations*, 2022.

[17] X. Chen and K. He, "Exploring simple Siamese representation learning," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15750–15758.

[18] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 15979–15988.

[19] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. International Conference on Learning Representations*, 2021.

[20] Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022.

[21] A. Dai, J. Yang, T. Zhang, B. Gao, K. Tang, and X. Chen, "DADR-HCD: A deep domain adaptation and disentangled representation network for unsupervised heterogeneous change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024, doi: 10.1109/TGRS.2024.3403727.

[22] Z. Ren et al., "Self-supervised learning guided by SAR image factors for terrain classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–18, 2024, doi: 10.1109/TGRS.2024.3386963.

[23] L. Xie, S. You, A. Liu, Y. He, C. Huang, and J. Deng, "Mitigating data constraints in crop mapping: A self-supervised framework integrating adaptive clustering, graph convolution and global spatiotemporal attention," *International Journal of Applied Earth Observation and Geoinformation*, vol. 144, Art. no. 104951, 2025.

[24] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, 2015.

[25] M. T. Chiu et al., "Agriculture-Vision: A large aerial image database for agricultural pattern analysis," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2825–2835, doi: 10.1109/CVPR42600.2020.00290.

[26] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "BigEarthNet: A large-scale benchmark archive for remote sensing image understanding," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 5901–5904, doi: 10.1109/IGARSS.2019.8900532.

[27] A. Lee, G. Kim, S.-J. Hong, S.-W. Kim, and G. Kim, "Classification of dead cocoons using convolutional neural networks and machine learning methods," *IEEE Access*, vol. 11, pp. 137317–137327, 2023, doi: 10.1109/ACCESS.2023.3338540.

[28] D. Udayanga et al., "Dual-mode multispectral imaging system for food and agricultural product quality estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–12, 2024, doi: 10.1109/TIM.2024.3369129.

[29] T. B. Pun, A. Neupane, R. Koech, and K. J. Owen, "Detection and quantification of root-knot nematode eggs from tomato plants using image analysis," *IEEE Access*, vol. 10, pp. 123190–123204, 2022, doi: 10.1109/ACCESS.2022.3223707.

[30] A. Abdalla, M. Lyu, M. Belete, and H. Cen, "Color consistency of UAV imagery using multichannel CNN-based image-to-image regression and residual learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024, doi: 10.1109/TGRS.2024.3363651.

[31] C. Li, Y. He, and X. Li, "Feature attention parallel aggregation network for single image

haze removal," *IEEE Access*, vol. 10, pp. 15322–15335, 2022, doi: 10.1109/ACCESS.2022.3148167.

[32] Y. Chai, H. Yu, L. Xu, D. Li, and Y. Chen, "Deep learning algorithms for sonar imagery analysis and its application in aquaculture: A review," *IEEE Sensors Journal*, vol. 23, no. 23, pp. 28549–28563, 2023, doi: 10.1109/JSEN.2023.3324438.

[33] X. Shi, X. Chai, J. Xie, and T. Sun, "MC-GCN: A multi-scale contrastive graph convolutional network for unconstrained face recognition with image sets," *IEEE Transactions on Image Processing*, vol. 31, pp. 3046–3055, 2022, doi: 10.1109/TIP.2022.3163851.

[34] L. Zhang et al., "A multi-scale convolutional hybrid attention residual network for enhancing underwater image and identifying underwater multi-scene sea cucumber," *IEEE Robotics and Automation Letters*, vol. 9, no. 9, pp. 7397–7404, 2024, doi: 10.1109/LRA.2024.3426382.

[35] Dr. R. Deepa, Vijaya Bhaskar Sadu, Dr. A.Sivasamy (2024). Early prediction of cardiovascular disease using machine learning: Unveiling risk factors from health records. *AIPAdvances*,14(3), https://doi.org/10.1063/5.0191990.

[36] VB Sadu, RS Kumar, BS Kumar, T Kavitha, HK Chapala, MK Chakravarthi (2024), Evaluating Machine Learning Models for Multimodal Probability-Based Energy Forecasting. *Process Integration and Optimization for Sustainability*, vol(8), pp:1209-1222. https://doi.org/10.1007/s41660-024-00428-0.

[37] Vijaya Bhaskar Sadu, Bharathi Subramaniam, Rukmani Devi Sethuraman, Sudhakar Tummala, Beluguri Venkateswarlu, Jeevika Tharini Viswanathan (2026). Balancing accuracy and efficiency in lumbar spine image segmentation using multi-scale attention and residual learning, *Scientific Reports*, https://doi.org/10.1038/s41598-026-58394-1