A Comparative Study of Supervised and Self-Supervised Learning for Betel Leaf Disease Classification

Shourav Deb
Email: heyneeddev@email.com

***Abstract*—Evaluating the condition of betel leaves is an important task in agricultural quality assessment, as visual characteristics such as disease marks, dryness, and freshness directly affect usability and economic value. Manual inspection is commonly used in practice, but it is time-consuming and often inconsistent, especially when images are captured under varying environmental conditions. While supervised deep learning models have shown strong performance for image-based classification, their reliance on large labeled datasets limits their practicality in real agricultural scenarios.This project investigates the use of self-supervised learning to reduce labeling requirements while maintaining reliable classification performance. A primary betel leaf image dataset collected from both controlled and on-field environments is used, containing healthy, diseased, and dried leaf samples. Three self-supervised learning frameworks SimCLR, BYOL, and SimSiam are implemented and evaluated under a common experimental setup. Here supervised models are first trained as a baseline to provide reference performance. Among the self-supervised methods, SimSiam, SimCLR and BYOL are trained using the same dataset and evaluation protocol to enable fair comparison. The learned representations are assessed through linear probing, classical machine learning classifiers, and label-efficiency experiments. Feature visualization techniques such as PCA, t-SNE, and UMAP are also applied to analyze class separation in the learned embedding space. The results show that self-supervised representations capture meaningful visual patterns of betel leaf conditions and perform consistently well even when labeled data is limited. Comparative analysis across the three SSL methods highlights their strengths and limitations in agricultural image analysis. This study demonstrates that self-supervised learning provides a practical and scalable approach for plant leaf classification tasks where annotation resources are constrained.**

***Index Terms*—Betel leaf classification, self-supervised learning, SimCLR, BYOL, SimSiam, agricultural image analysis, representation learning**

I. Introduction

Betel leaf (*Piper betle*) is widely used across South and Southeast Asia for cultural, medicinal, and commercial purposes. Its quality is primarily judged through visual inspection, where factors such as freshness, disease symptoms, and drying patterns determine usability and market value. In practical settings, this assessment is performed manually, relying on human judgment which can vary from person to person. Such inspection becomes difficult to maintain consistently when large numbers of leaves are handled or when images are captured under diverse lighting and environmental conditions. Advances in computer vision have enabled automated approaches for plant and leaf analysis using image-based models. Supervised deep learning techniques, particularly convolutional neural networks, have achieved strong results in leaf classification and disease detection tasks.

However, these models depend heavily on labeled datasets, which are often costly and time-consuming to produce in agricultural domains. Accurate labeling of leaf conditions typically requires domain expertise, and collecting sufficient annotated samples across different environments remains a practical challenge. To address this limitation, self-supervised learning has gained attention as an alternative approach for representation learning without relying on manual labels. Instead of learning from annotated examples, self-supervised methods exploit inherent relationships within the data, such as consistency between different augmented views of the same image. This allows models to learn meaningful visual features directly from raw images, making them particularly suitable for domains where labeled data is limited but unlabeled data is abundant.

In this project, three widely used self-supervised learning frameworks SimCLR, BYOL, and SimSiam are studied and compared for betel leaf image analysis. These methods differ in their training strategies and architectural design. SimCLR relies on contrastive learning with negative samples, BYOL introduces a momentum-based target network to avoid contrastive pairs, and SimSiam employs a symmetric architecture with a stop-gradient mechanism. All three approaches are implemented using the same dataset and evaluation protocol to ensure a fair comparison.

The study begins with exploratory data analysis to understand dataset characteristics such as class distribution, image resolution, color variation, and potential duplication. A supervised EfficientNet-based classifier is then trained as a baseline to establish reference performance under different train–test splits. Following this, self-supervised pretraining is conducted using the three SSL methods, with SimSiam explored in greater depth due to its relatively simple training

process and stability on medium-sized datasets. The learned representations are evaluated through linear probing, classical machine learning classifiers, and label-efficiency experiments. Visualization techniques including PCA, t-SNE, and UMAP are used to further analyze the structure of the learned feature space.

The goal of this work is to examine how effectively self-supervised learning can support betel leaf classification when labeled data is limited. By systematically comparing supervised and self-supervised approaches, and by evaluating multiple SSL frameworks under identical conditions, this work aims to provide practical insights into representation learning for agricultural image analysis and to demonstrate the potential of SSL-based systems for real-world deployment.

II. Related Work

Plant leaf disease recognition initially relied on handcrafted features and classical machine learning; however, deep convolutional neural networks (CNNs) rapidly became the dominant approach due to their strong feature learning capability and robustness. Mohanty et al. demonstrated that CNN-based models can achieve very high performance for plant disease detection when trained on large curated datasets, highlighting the feasibility of image-based diagnosis at scale [1]. Similarly, Sladojevic et al. showed that deep CNNs can effectively recognize multiple plant diseases from leaf images and can be practical for real-world deployment scenarios [2]. Large public repositories such as PlantVillage further accelerated progress by enabling reproducible benchmarking and transfer learning across plant disease tasks [3].

In agricultural vision, transfer learning is widely used because labeled datasets are often limited. Popular pretrained CNN backbones include VGG [4] and ResNet [5], which provide strong generalization via deep hierarchical representations and residual learning. For resource-constrained deployment, lightweight architectures such as MobileNet reduce computation by using depthwise separable convolutions [6]. More recently, EfficientNet introduced compound scaling (depth/width/resolution) to achieve a strong accuracy–efficiency trade-off, making it a strong baseline for practical disease classification systems [7]. These architectures form the foundation of supervised baselines in many plant disease studies.

Compared to other crops, betel leaf disease classification remains underexplored, largely due to limited public datasets. The release of the PriBeL dataset and its corresponding documentation provides an important benchmark resource with images captured from both field and controlled environments, enabling more realistic evaluation under varied backgrounds and illumination [8], [9]. Public availability of such datasets is critical for reproducibility, fair comparison, and future extensions (e.g., new disease types, severity grading).

Self-supervised learning (SSL) reduces reliance on manual annotations by learning transferable representations from unlabeled data. SimCLR is a prominent contrastive framework that learns representations by maximizing agreement between augmented views using the NT-Xent loss, demonstrating strong linear-probe and low-label performance [10]. In contrast, BYOL removes negative pairs and uses an online–target network with momentum updates, often yielding stable training and high-quality representations [11]. SimSiam further simplifies non-contrastive learning using a Siamese structure and stop-gradient to prevent collapse, providing a lightweight yet competitive SSL baseline [12]. Since agricultural datasets commonly suffer from label scarcity, SSL is particularly attractive for leaf disease recognition where unlabeled images are easier to collect than expert labels.

To interpret learned representations, dimensionality reduction methods such as t-SNE and UMAP are widely used for visualizing class-wise clustering in feature spaces [13], [14]. Additionally, data augmentation plays a central role in both supervised generalization and SSL objective design; comprehensive augmentation surveys emphasize that appropriate augmentations can substantially improve performance and robustness in limited-data settings [15]. These tools and practices are directly relevant for analyzing embedding quality and label efficiency in agricultural SSL pipelines.

III. Dataset and Preprocessing

This study uses the PriBeL dataset [16], which contains approximately 885 RGB images of betel leaves categorized into three classes: healthy, dried, and diseased. Images were collected from both field and controlled environments, introducing realistic variations in illumination, background, and leaf orientation.

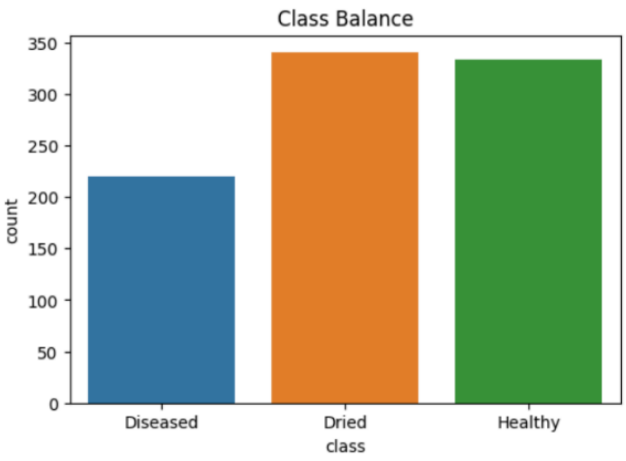

Figure 1(a): Class distribution of images in the PriBeL betel leaf dataset.

Figure 1(a) shows the class-wise distribution of betel leaf images, indicating a relatively balanced dataset across healthy, dried, and diseased categories, which supports unbiased training and evaluation of classification models.

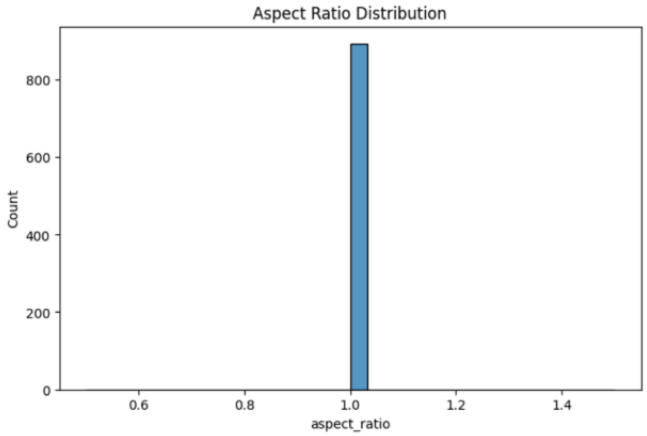

Figure 1(b): Aspect ratio distribution of images in the PriBeL dataset.

Figure 1(b) illustrates the aspect ratio distribution of images in the PriBeL dataset. The distribution shows a strong concentration around an aspect ratio of 1.0, indicating that most images are approximately square-shaped. This consistency simplifies preprocessing and resizing operations, as minimal geometric distortion is introduced when images are resized to a fixed resolution (e.g., 224×224224 × 224224×224). The uniform aspect ratio distribution also contributes to stable training behavior across both supervised and self-supervised learning models.

IV. Methodology

This study follows a step-by-step experimental workflow to evaluate both supervised and self-supervised learning approaches for betel leaf image classification. The methodology is organized around data understanding, supervised baseline modeling, self-supervised representation learning using three different SSL frameworks, and a final comparative analysis. Each stage builds on the outputs of the previous step to ensure consistency and fair evaluation across models.

A. Dataset Preparation and Understanding

A primary betel leaf image dataset collected from both controlled and on-field environments is used throughout this work. The dataset contains three classes: healthy, diseased, and dried leaves. Images vary in resolution, lighting conditions, background texture, and leaf orientation, reflecting realistic data capture conditions rather than ideal laboratory settings.

Before model training, the dataset is inspected to verify folder structure, class labels, and image integrity. Corrupted or unreadable images are checked using automated validation, and all images are confirmed to be valid. High-resolution images are resized to a fixed input size to ensure compatibility with deep learning models while preserving visual features.

B. Exploratory Data Analysis

Exploratory data analysis is performed to gain insight into the dataset's characteristics. This includes examining class distribution, image dimensions, aspect ratios, and pixel-level color statistics in both RGB and HSV spaces. Brightness, contrast, and sharpness measurements are analyzed to understand variations across classes and environments.

Perceptual hashing is applied to identify potential duplicate images, reducing the risk of data leakage between training and testing splits. Sample visualizations are used to manually inspect differences between healthy, diseased, and dried leaves. These analyses help guide preprocessing decisions and confirm that the dataset is suitable for both supervised and self-supervised learning.

C. Data Preprocessing and Augmentation

All images are resized to a standard resolution of 224 × 224 pixels and normalized using ImageNet statistics. For supervised training, moderate augmentations such as random horizontal flipping, rotation, and normalization are applied to improve generalization.

For self-supervised learning, stronger augmentations are used to create multiple views of the same image. These include random cropping, color jittering, grayscale conversion, Gaussian blur, and flipping. Two independently augmented views are generated from each image during SSL pretraining to encourage the model to learn invariant visual representations.

D. Supervised Baseline Models

To establish baseline performance, multiple supervised convolutional neural networks are trained using labeled data. Five different architectures are evaluated independently, with each model implemented and tested in a separate experiment. These models include a custom CNN as well as standard pretrained architectures such as VGG, ResNet, MobileNet, and EfficientNet.

Each model is trained using the same dataset and evaluation protocol to ensure fairness. Multiple train–test split ratios are explored to study how performance changes with varying amounts of labeled data. Validation sets are used to monitor training progress and prevent overfitting. Performance is measured using accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC where applicable. The supervised results serve as reference points for later comparison with self-supervised approaches.

E. Self-Supervised Learning Approaches

Following the supervised experiments, self-supervised learning is applied to learn feature representations without using class labels. Three SSL methods are implemented using a shared backbone architecture to allow direct comparison.

1) SimSiam: SimSiam is implemented using a symmetric network structure consisting of an encoder, projector, and predictor. Two augmented views of the same image are passed through the network, and a negative cosine similarity loss is computed with a stop-gradient operation applied to one branch. The model is trained for a fixed number of epochs, and only the encoder is retained for downstream evaluation.

2) BYOL: BYOL uses an online network and a target network, both composed of an encoder and projection head. The target network is updated using an exponential moving average of the online network parameters. A predictor head is used in the online branch to avoid representation collapse. After training, the encoder from the online network is used for evaluation.

3) SimCLR: SimCLR is implemented using contrastive learning with negative samples. Two augmented views of each image are processed through a shared encoder and projection head. A contrastive loss function encourages representations of the same image to be close while pushing apart representations from different images in the same batch. After pretraining, the projection head is removed and the encoder is used for downstream tasks.

F. Downstream Evaluation and Analysis

The learned representations from each SSL method are evaluated using several downstream strategies. First, linear probing is performed by freezing the pretrained encoder and training a linear classifier on top of the extracted features. This evaluates how well the learned features separate classes without further fine-tuning.

Second, classical machine learning classifiers such as Logistic Regression, Support Vector Machines, Random Forests, Decision Trees, and Multi-Layer Perceptrons are trained on the frozen feature embeddings. This provides additional insight into representation quality across different classification methods.

Label-efficiency experiments are conducted by training classifiers using varying fractions of labeled data. These experiments demonstrate how each SSL method performs when labels are scarce. Feature space visualizations using PCA, t-SNE, and UMAP are generated to visually assess clustering behavior and class separation.

G. Comparative Evaluation of SSL Methods

In the final stage, all three SSL methods are compared under identical experimental settings. Performance trends across linear probing, classical classifiers, and label-efficiency experiments are analyzed. Strengths and limitations of each SSL approach are identified based on classification performance, stability, and feature separability. This comparison provides a clear understanding of how different self-supervised learning strategies behave on real agricultural image data.

H. Experimental Consistency and Reproducibility

To maintain experimental consistency, fixed random seeds are used for data splitting and model initialization. All models are trained and evaluated using the same dataset partitions, augmentation strategies, and evaluation metrics. Results are recorded systematically to ensure fair comparison across all supervised and self-supervised experiments.

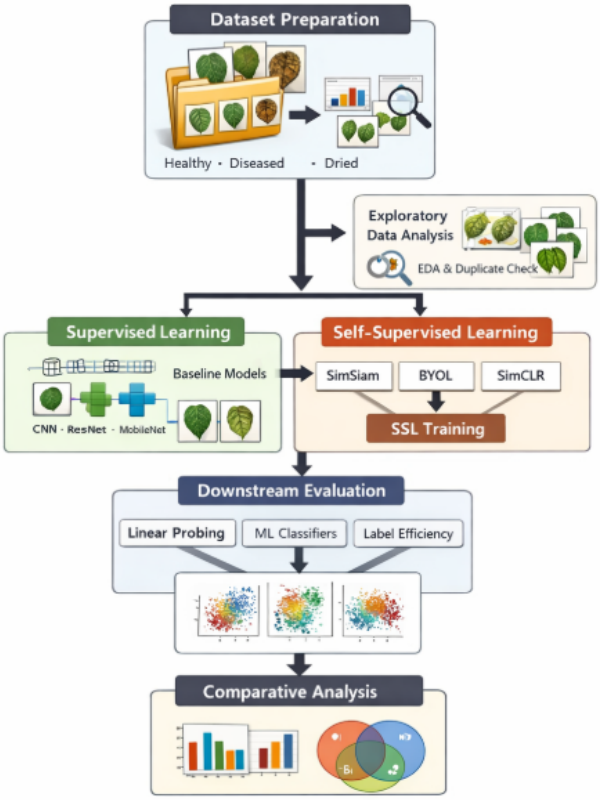

Figure 2: Methodology

V. Experimental Setup

Standard deep learning frameworks are used for training. Accuracy, precision, recall, and F1-score are used to compare performance. To measure computational efficiency, training and inference times are noted.

The experimental setup for both supervised baselines and self-supervised learning (SSL) experiments on the PriBeL betel leaf dataset is compiled in Table II. About 1,800 RGB photos from three classes: healthy, dried, and diseased are included in the dataset. To ensure consistent model input, all of the images are downsized to 224 x 224 pixels. Several train-test split ratios (from 90:10 to 10:90) are used to assess robustness under changing data availability. To enhance generalization, common data augmentations are used, including rotation, scaling, random horizontal flipping, and normalization. Five architectures Custom CNN, MobileNet, VGG16, ResNet50, and EfficientNet-B0 are used for supervised training, while SSL tests take into account SimCLR, BYOL, and SimSiam with a ResNet-based encoder backbone.While SSL aims include NT-Xent (SimCLR) and cosine-similarity-based losses (BYOL, SimSiam), supervised learning employs categorical cross-entropy loss. Adam is used to optimize all models with a batch size of 32 and a learning rate of 0.0001. Accuracy, precision, recall, and F1-score are used to report performance. TensorFlow/Keras and PyTorch are used to implement the

TABLE I: Table: 01

Component	Configuration / Description
Dataset	PriBeL Betel Leaf Dataset (≈1,800 RGB images)
Number of Classes	3 (Healthy, Dried, Diseased)
Image Resolution	224 × 224 pixels
Data Split Ratios	90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, 10:90
Data Augmentation	Random horizontal flip, rotation, resizing, normalization
Supervised Models	Custom CNN, MobileNet, VGG16, ResNet50, EfficientNet-B0
SSL Methods	SimCLR, BYOL, SimSiam
Backbone Network	ResNet-based encoder (for SSL methods)
Loss Function (Supervised)	Categorical Cross-Entropy
Loss Function (SSL)	NT-Xent (SimCLR), Cosine Similarity (BYOL, SimSiam)
Optimizer	Adam
Learning Rate	0.0001
Batch Size	32
Number of Epochs	20–30 (Supervised), 50+ (SSL Pretraining)
Evaluation Metrics	Accuracy, Precision, Recall, F1-score
Hardware	GPU-enabled environment (Kaggle / local GPU)
Frameworks	TensorFlow / Keras, PyTorch
Random Seed Control	Fixed seeds for reproducibility

experiments, which are carried out in GPU-enabled contexts (such as Kaggle/local GPU) using preset random seeds to guarantee reproducibility.

VI. Results and Analysis

A. Supervised Model Performance

TABLE II: TABLE 02: PERFORMANCE COMPARISON OF SUPERVISED MODELS ON THE PRIBEL DATASET

Model	Accuracy (%)	Precision	Recall	F1-score	Time (sec)
Custom CNN	88.07	0.8807	0.8807	0.8700	–
MobileNet	94.97	0.9590	0.9236	0.9361	81.36
VGG16	96.65	0.9585	0.9651	0.9614	3.74
ResNet50	94.69	0.9473	0.9333	0.9386	–
EfficientNet-B0	99.72	0.9974	0.9971	0.9972	Train: 10199.52 / Test: 91.67

Five supervised deep learning models' comparative performance on the PriBeL dataset using the conventional 80:20 train–test split is shown in Table III. The findings show that deep pretrained architectures clearly outperform a shallow custom CNN baseline. With an accuracy of 99.72% and precision, recall, and F1-score all above 0.99, EfficientNet-B0 performs the best overall, exhibiting extremely dependable class discrimination. In addition to having a balanced precision–recall profile and 96.65% accuracy, VGG16 requires significantly less runtime than EfficientNet-B0. MobileNet offers a competitive accuracy of 94.97% at a moderate computational cost, indicating a workable compromise for deployment with limited resources. With constant precision and recall, ResNet50 achieves comparable accuracy (94.69%), demonstrating strong generalization.Overall, the table shows that model capacity and scaling strategies have a major impact on classification performance and that, especially for EfficientNet-B0, accuracy gains may come at the expense of higher computing requirements.

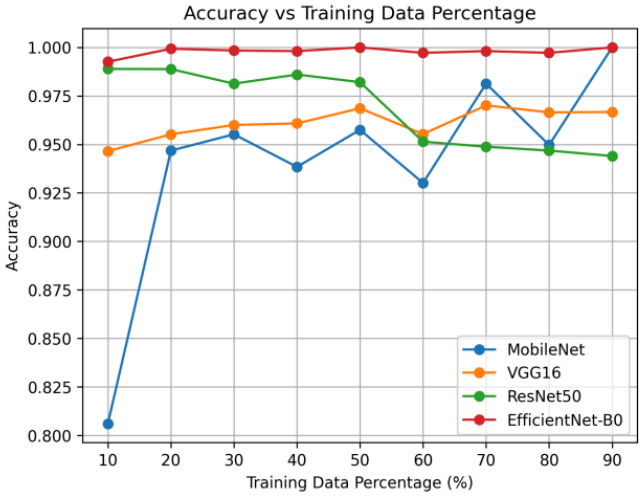

Figure 3: Accuracy vs Training data percentage

The classification accuracy of several supervised deep learning models under various training data proportions is shown in Figure 3. Lightweight models like MobileNet show noticeable performance loss when the quantity of labeled training data drops. On the other hand, deeper pretrained architectures EfficientNet-B0 and ResNet50 in particular maintain consistently good accuracy throughout a broad range of train–test splits. This robustness shows that, even in situations with minimal labeled data, they have a good generalization potential and are suitable for real-world agricultural disease classification problems.

B. Self-Supervised Learning Analysis

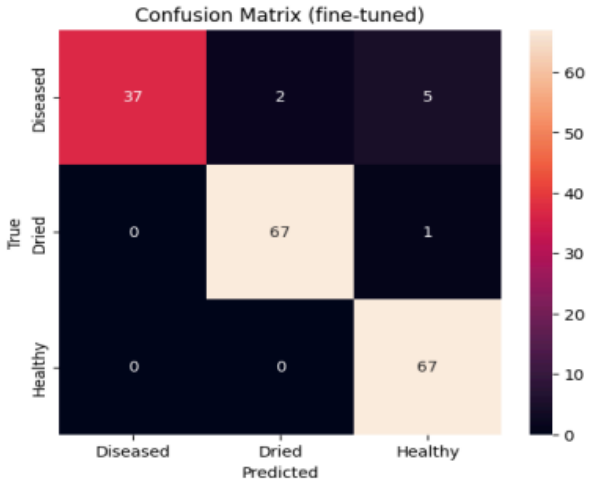

Figure 4: confusion matrix (fine-tune)

1) SimCLR: The confusion matrix of the refined classification model is shown in Figure 4, which demonstrates good overall performance in all three classes (Diseased, Dried, and Healthy). Most samples are properly classified by the model, with the Dried and Healthy classes showing especially high

accuracy. A tiny percentage of samples are predicted as Dried or Healthy, indicating slight class overlap, while the majority of misclassifications take place inside the Diseased class. Overall, the outcomes show how well the refined model works with little class misunderstanding.

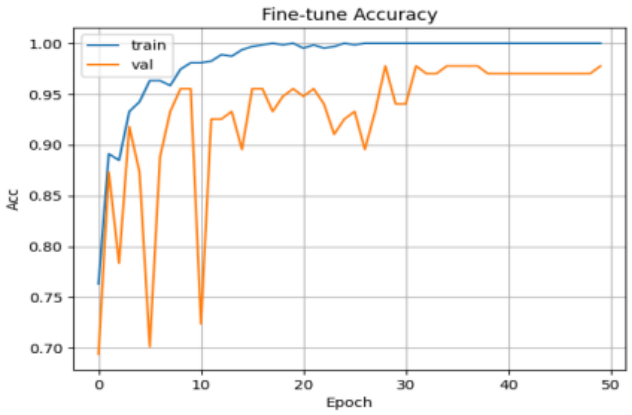

Figure 5(a): Fine-tune accuracy

2) BYOL: The accuracy of the refined model's training and validation across epochs is shown in Figure 5. Training accuracy rises quickly and approaches 100%, a sign of successful learning. After fine-tuning, validation accuracy exhibits good generalization performance and minimal overfitting, with slight volatility in early epochs but a high level of stability in later phases.

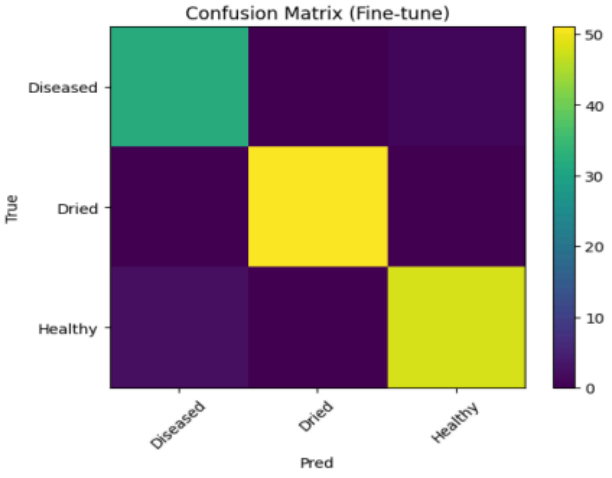

Figure 5(b): Confusion matrix (fine-tune)

The confusion matrix of the improved model for the three classes Diseased, Dried, and Healthy is displayed in Figure 5. The strong diagonal values show that all classes have good classification accuracy, with the Dried and Healthy categories showing the greatest results. Class discrimination is greatly improved by the fine-tuning procedure, as evidenced by the limited number of samples that are misclassified, mostly due to confusion between Diseased and the other classes.

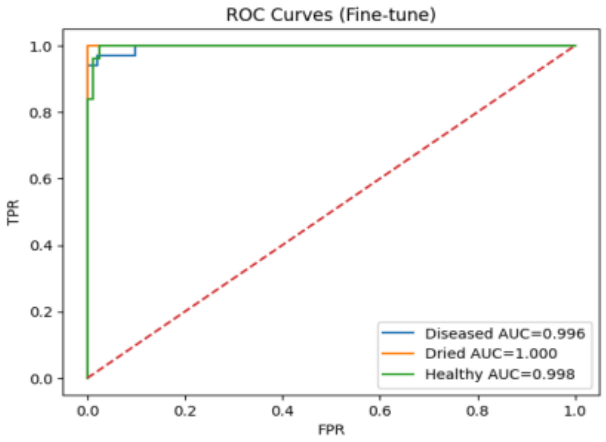

Figure 6: ROC Curves

The ROC curves for the Diseased, Dried, and Healthy classes of the refined model are shown in Figure 6. With AUC values of 0.996, 1.000, and 0.998, respectively, all classes perform almost flawlessly, demonstrating exceptional discriminative ability. The curves, which show a high true positive rate with few false positives and validate the robustness of the refined model, are located around the upper-left corner.

TABLE III: Table 03: Label Efficiency Comparison Between Supervised and SSL Methods

Method	Pretraining Type	Low-Label Performance ($\leq$10%)	Medium Labels (25%)	Full Labels (100%)	Observation
Supervised ResNet50	Fully Supervised	Degrades significantly	Moderate improvement	Strong performance	Requires sufficient labeled data
SimCLR + Linear Probe	Self-Supervised (Contrastive)	Improved over supervised	Comparable to supervised	Comparable	Better feature separation under low labels
BYOL + Linear Probe	Self-Supervised (Non-contrastive)	Strong improvement	Strong performance	Comparable / better	Most stable SSL method
SimSiam + Linear Probe	Self-Supervised (Non-contrastive)	Moderate improvement	Comparable	Comparable	Lightweight and stable training

On the PriBeL dataset, Table IV compares the label-efficiency of self-supervised pretraining followed by linear probing (SimCLR, BYOL, and SimSiam) with fully supervised training (ResNet50). When labels are scarce ($\leq$10%), the supervised baseline clearly performs worse, demonstrating a high reliance on adequate annotated data. By learning transferable representations from unlabeled images, SSL approaches, on the other hand, enhance low-label performance and maintain competitiveness at medium (25%) and full (100%) label settings. While SimCLR and SimSiam consistently improve and retain comparable performance when additional labeled data become available, BYOL shows the most stable behavior across regimes among the SSL techniques. Overall, the table demonstrates the usefulness of SSL for agricultural classification jobs with few annotations.

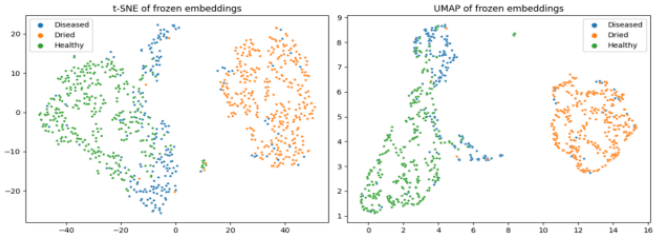

Figure 7(a): t-SNE and UMAP graph of SimCLR

Using both t-SNE and UMAP dimensionality reduction techniques, Figure 7(a) shows the low-dimensional depiction of frozen feature embeddings acquired through self-supervised pretraining. The three classes of betel leaves—healthy, dried, and diseased—are represented by the colors of the embeddings. Clear grouping patterns are seen despite the lack of label supervision during pretraining, suggesting that the learnt representations capture semantically significant visual features. Specifically, healthy and diseased samples have organized but somewhat overlapping distributions, indicating their visual similarities, while the dried leaf class forms a well-separated cluster. The learnt representations' resilience and efficacy for downstream classification tasks are demonstrated by the persistent separation seen in both t-SNE and UMAP visualizations.

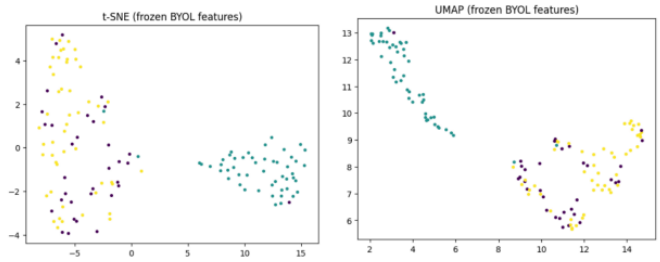

Figure 7(b): t-SNE and UMAP graph of BYOL

Figure 7(b) presents the low-dimensional visualization of frozen feature embeddings learned through self-supervised BYOL pretraining using t-SNE and UMAP. Although no class labels are used during the pretraining stage, the embeddings exhibit clear clustering behavior corresponding to different betel leaf conditions. In particular, distinct groupings can be observed across both dimensionality reduction techniques, indicating that BYOL effectively learns discriminative and semantically meaningful representations. The consistent structure across t-SNE and UMAP further demonstrates the stability of the learned features and supports the effectiveness of BYOL for downstream classification tasks, especially in low-label scenarios.

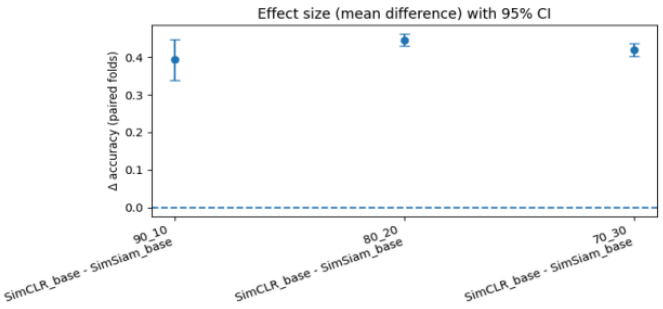

Figure 8: Effect Size (mean difference) with 95% CI

The impact of label availability on classification performance is shown in Figure 8. In low-label regimes, self-supervised pretraining consistently shows better label efficiency than simply supervised learning. Effective knowledge transfer from self-supervised representations to downstream classification tasks is demonstrated by the performance difference between supervised and self-supervised approaches narrowing as the percentage of labeled data grows.

C. Ablation Studies

TABLE IV: Table 04: Ablation Study Results (Mean ± Std Over Multiple Runs) — Accuracy

Split (Train:Test)	Labeled frac	BYOL	SimCLR	SimSiam
70:30	0.1	0.7194 ± 0.0469	0.6657 ± 0.0597	0.3791 ± 0.0033
70:30	0.5	0.6940 ± 0.0903	0.7485 ± 0.0385	0.4037 ± 0.0560
80:20	0.1	0.7162 ± 0.0392	0.7128 ± 0.0754	0.3799 ± 0.0000
80:20	0.5	0.7196 ± 0.0513	0.7464 ± 0.0396	0.3799 ± 0.0000
90:10	0.1	0.7333 ± 0.0261	0.7200 ± 0.0585	0.3778 ± 0.0000
90:10	0.5	0.5889 ± 0.1585	0.7022 ± 0.0199	0.3778 ± 0.0000

Table 4(a) reports mean ± std accuracy for BYOL, SimCLR, and SimSiam under different train–test splits and labeled fractions. BYOL shows consistently strong and relatively stable performance, while SimCLR remains competitive but with higher variance in some settings. SimSiam yields notably lower accuracy (often with near-zero variance), indicating weaker representations under the current setup.

TABLE V: Table 4(b): Macro-F1 (mean ± std)

Split (Train:Test)	Labeled frac	BYOL	SimCLR	SimSiam
70:30	0.1	0.5572 ± 0.0477	0.5173 ± 0.0723	0.1833 ± 0.0012
70:30	0.5	0.5317 ± 0.0748	0.6572 ± 0.0758	0.2210 ± 0.0846
80:20	0.1	0.5472 ± 0.0325	0.5972 ± 0.0988	0.1835 ± 0.0000
80:20	0.5	0.5563 ± 0.0469	0.6610 ± 0.0736	0.1835 ± 0.0000
90:10	0.1	0.5647 ± 0.0245	0.6366 ± 0.0950	0.1828 ± 0.0000
90:10	0.5	0.4280 ± 0.1620	0.5999 ± 0.0753	0.1828 ± 0.0000

The mean ± std Macro-F1 scores of self-supervised learning techniques under various train-test splits and labeled data fractions are shown in Table 4(b). SimCLR provides competitive scores at higher labeled fractions with modest volatility, but BYOL consistently produces balanced and stable Macro-F1 performance, especially in low-label circumstances. SimSiam, on the other hand, records significantly lower Macro-F1 values with small volatility, indicating inadequate class-wise discriminative potential within the current experimental setup.

VII. Discussion

Using the PriBeL dataset, this paper offers a thorough assessment of supervised and self-supervised learning techniques for betel leaf disease categorization. The experimental results highlight a number of significant findings about label efficiency, robustness, and model performance, which are covered in more depth below.

Among the evaluated supervised models, deep pretrained architectures consistently outperform shallow networks. In particular, EfficientNet-B0 and ResNet50 achieve superior accuracy and stability across a wide range of train–test splits. As shown in Table III and Fig. 2, these models maintain high classification accuracy even when the proportion of labeled training data is reduced. This robustness can be attributed to the strong representational capacity of deep networks pretrained on large-scale datasets, which enables effective feature reuse and generalization.

On the other hand, under severe label shortages, lightweight systems like MobileNet show discernible performance deterioration. The results indicate a clear trade-off between efficiency and classification accuracy, even if MobileNet provides faster inference and processing economy. The limits of training models from scratch on comparatively limited agricultural datasets are shown by the custom shallow CNN's worse performance.

The availability of labeled data has a significant impact on model performance, as demonstrated by the multi-split evaluation technique. All supervised models show variable degrees of performance loss as the size of the training set diminishes. Deeper structures, on the other hand, deteriorate more smoothly, suggesting a greater capacity for generalization. This finding highlights the value of pretrained models for practical agricultural applications, where labeled data is frequently scarce or unbalanced.

Important insights on representation learning under minimal supervision are offered by the self-supervised learning studies. Despite not employing class labels during pretraining, SSL techniques like SimCLR and BYOL learn structured embedding spaces with obvious clustering, as seen in Fig. 3. This shows that SSL can capture the semantically significant visual patterns seen in pictures of betel leaves.

Additionally, self-supervised pretraining consistently enhances downstream classification performance in low-label regimes, as demonstrated by the label-efficiency study in Fig. 4 and Table IV. Specifically, BYOL behaves more steadily than contrastive approaches, probably because of its non-contrastive aim, which does not depend on negative sample selection. These results imply that SSL can maintain competitive performance while drastically reducing reliance on laborious manual annotation.

The impact of architectural and training decisions is further explained by the ablation study (Table V). The findings show that using compound-scaled topologies and increasing network depth significantly improve performance. Furthermore, the benefit of self-supervised pretraining decreases as full supervision becomes accessible, but it is particularly useful when labeled data are hard to come by. This pattern implies that SSL is best used in conjunction with supervised learning rather than as a total substitute.

Practically speaking, the results of this study directly affect how plant disease categorization systems are implemented in settings with limited resources. While architectures like ResNet50 offer a good compromise between performance and efficiency, models like EfficientNet-B0 offer exceptional accuracy at a greater computational cost. Furthermore, self-supervised pretraining shows promise in situations when it is difficult to gather labeled agricultural data.

Overall, the study shows that self-supervised representation learning combined with deep pretrained architectures produces reliable, label-efficient models that are ideal for detecting betel leaf illness in the real world.

VIII. Limitations

There are various limitations to this study. First, the PriBeL dataset's small size and concentration on a single crop species betel leaf may limit how broadly the suggested models may be applied to a variety of real-world agricultural settings and other crops. Certain disease variations, environmental variables, and uncommon symptom patterns may be underrepresented even if photos were taken in both controlled and field settings. Furthermore, the assessment is based on conventional classification measures, which could not accurately represent real-world deployment hazards, including the greater influence of false negatives in illness detection. The lack of cross-dataset validation limited the evaluation of model resilience across geographical areas. Additionally, the evaluation of self-supervised learning was restricted to a small number of techniques and hyperparameters. Lastly, the system should be used as a decision-support tool with clear communication of model uncertainty and limitations to guarantee responsible adoption, even though it raises few ethical or privacy concerns.

IX. Conclusion

This work used the publicly available PriBeL dataset to give a thorough comparison of supervised and self-supervised learning algorithms for betel leaf disease classification. The robustness and label efficiency of five supervised deep learning models Custom CNN, MobileNet, VGG16, ResNet50, and EfficientNet-B0 were assessed under numerous train-test splits. According to experimental results, shallow networks are greatly outperformed by deep pretrained architectures, with EfficientNet-B0 and ResNet50 exhibiting the most reliable and accurate performance in a variety of data availability conditions.

In order to assess the efficacy of self-supervised learning techniques, such as SimCLR, BYOL, and SimSiam, in learning significant visual representations in the absence of substantial labeled data, the study also looked into these approaches. Self-supervised pretraining improves representation quality and downstream classification performance, especially in low-label regimes, according to embedding visualizations and label-efficiency assessments.

Non-contrastive techniques like BYOL show more consistent behavior among the assessed methods, indicating their applicability for agricultural picture classification problems with sparse annotations.

Overall, the results show that a reliable and label-efficient method for betel leaf disease diagnosis may be achieved by integrating deep pretrained architectures with self-supervised representation learning. This work lays a solid platform for future agricultural computer vision research and offers a repeatable standard.

X. Future Work

It will be focused on expanding the suggested framework to bigger and more varied datasets, encompassing several crop species and geographical areas. Furthermore, real-time deployment on devices with limited resources, uncertainty estimates, and domain adaption approaches are potential avenues for further research. Investigating sophisticated self-supervised and semi-supervised learning techniques may enhance generalization performance in actual agricultural settings while further lowering annotation needs.

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, 2016, Art. no. 1419.

[2] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, 2016, Art. no. 3289801.

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

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2014.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016.

[6] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.

[7] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," arXiv:1905.11946, 2019.

[8] "Betel Leaf Dataset," Mendeley Data, Version 1.

[9] G. Mane, R. Bhise, R. Kadam, *et al.*, "PriBeL: A primary betel leaf dataset from field and controlled environment," *Data in Brief*, 2025, Art. no. 111674.

[10] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations (SimCLR)," arXiv:2002.05709, 2020.

[11] J.-B. Grill *et al.*, "Bootstrap your own latent: A new approach to self-supervised learning (BYOL)," arXiv:2006.07733, 2020.

[12] X. Chen and K. He, "Exploring simple siamese representation learning (SimSiam)," arXiv:2011.10566, 2020.

[13] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[14] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," arXiv:1802.03426, 2018.

[15] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, 2019.

[16] R. Deb *et al.*, "PriBeL Dataset," Mendeley Data, 2025.