Self-Supervised Learning for Class-Imbalanced Guava Disease Detection: A Contrastive Reconstruction Enhancement Approach

MSc Research Project
Master's in Data Analytics

Vishnu Vardhan Reddy Rikkula
Student ID: X23409452

School of Computing
National College of Ireland

Supervisor: Mr. Jaswinder Singh

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Vishnu Vardhan Reddy Rikkula

Student ID: X23409452

Programme: Master's in Data Analytics **Year:** 2025-26

Module: Research Practicum

Supervisor: Mr. Jaswinder Singh

Submission Due Date: 28-01-2026.

Project Title: Self-Supervised Learning for Class-Imbalanced Guava Disease Detection: A Contrastive Reconstruction Enhancement Approach.

Word Count:9628 **Page Count: 23**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Vishnu Vardhan Reddy Rikkula

Date: 28-01-2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Self-Supervised Learning for Class-Imbalanced Guava Disease Detection: A Contrastive Reconstruction Enhancement Approach

Vishnu Vardhan Reddy Rikkula
X23409452

Abstract
Supervised learning based systems of plant disease detection are highly accurate in general, but fail utterly on off-the-record disease classes. This renders them inefficient in agriculture where isolated diseases tend to go viral. This research paper discusses the effectiveness of self-supervised learning in alleviating serious class imbalance in guava disease detection with specific attention to the pre-training of Contrastive Reconstruction Enhancement (CRE) in minority class detection in comparison to full supervision methods. The experimental design applied in the study is the three-phase one, that is, setting supervised baselines on the Bangladesh guava dataset (4,360 images, 6 classes), running the CRE pre-training procedure on the PlantVillage dataset (61,486 images, 39 disease classes), and conducting comparative analysis through fine-tuning. With the outcomes of the experiment, it is evident that CRE pre-training significantly enhances the identification of minority classes. As an example, Healthy leaves increased in the recall percentage, 67% to 81.82% and Insects eaten increased in the recall percentage, 50% to 76.92%. The macro F1 score increased to 0.9406 as compared to 0.8930 and this indicates that there was an enhancement of performance in all the classes by 5.33%. Although the total accuracy in this case decreased a little bit, setting 98.77 % at 97.71%, this trade-off is favorable to practical application when it is of great interest to find a rare disease. The paper contains some empirically proven evidence that class imbalance in the agricultural AI could be addressed through self-supervised learning, which would be a replicated methodology applicable in the situation with other special crops faced with similar data constraints. These findings indicate that the use of SSL is an effective method of developing powerful disease detection systems that may be employed in agricultural contexts where available resources are scarce.

1 Introduction

Plant diseases are becoming a growing menace to agricultural sector of Bangladesh and putting millions of food security and livelihoods of rural population at risk. With a load of 9 percent of total fruit production in the country, a guava of 12,250 hectares suffers losses of 15-20 percent or a year due to disease infection (Shihab et al., 2025). Such losses have a direct bearing of economic sustainability in farming communities, especially smallholder farmers who do not have resources to manage the disease comprehensively. Conventional methods of detection involve use of human eyes of the agricultural professionals, which is time consuming, subjective and unavoidably unavailable in the remote rural regions where majority of the farming is done.

The latest developments in deep learning have shown a possible high automated detection of plant disease with great accuracy on benchmark datasets. But there is one key obstacle to the application of agriculture: harsh class imbalance in practice in disease distributions. The prevalent diseases characterize the training data but infrequent diseases that have huge economic impacts are underrepresented. This asymmetry gives rise to a paradox in which the models are rated to have great overall accuracy, but completely fail in cases of minority disease classes that need to be identified quickly to avoid spreading the epidemics.

This is a critical challenge as is the case of the Bangladesh guava disease data. The dataset has 606 original images shared by six disease categories; therefore it has a very severe imbalance with minority classes of less than 4 per cent of the total number of samples. To be more specific, Healthy leaves has 150 and Insects eaten 164 samples as compared to majority classes such as Algal spot that has more than 1,300 images. The original experiments on state-of-the-art supervised learning architectures were first indicative of this essential drawback: EfficientNet V2-S can attain 98.77% test accuracy, but also exhibits catastrophic failure on the minority classes whose recall is 50% on Insects eaten, and 67 on Healthy leaves. Such systems are not suitable to the real life application where such performance fails to detect rare diseases thus leading to quick contamination of the pathogen and major losses of funds.

These constraints of the approach of supervised learning require the need to seek other alternatives that can adequately leverage the limited labelled data with even performance across all disease types. Self-supervised learning (SSL) has become an interesting paradigm that initially learns generalizable features representations by large unlabeled datasets and then fine-tunes on downstream tasks. This method can be especially applicable to agricultural use cases where it is still very cheap to acquire unlabeled images but is costly and time-intensive to have images annotated by an expert. CRE is one of the SSL approaches; it provides a hybrid structure of masked image modelling to understand the local disease symptoms and contrastive learning to enforce global semantic consistency. This two-fold aim fits well along the lines of plant disease characteristics that appear as localized lesions as well as the general change in appearance of the leaf as a whole.

Although the concept of the use of SSL is promising, its role in the task of class imbalance that is severe when dealing with specialized crop diseases has not yet been empirically confirmed. The current body of research on agricultural AI is mostly on typical crops but one which have a large amount of data and there is no thorough study on special crops like the guava in South Asian setting. This discrepancy encourages exploring how this fact may be addressed by means of having Multi-label learning surmount the main disadvantage of supervised learning: the absolute failure on minority classes of diseases even though it has a strong overall accuracy.

Research Question: How does self-supervised pre-training using Contrastive Reconstruction Enhancement affect the detection of minority disease classes compared to standard supervised learning approaches in guava disease classification?

Research Objectives:

1. **Develop a baseline performance standard** by assessing state-of-the-art monitored learning designs at the imbalanced Bangladesh guava disease dataset, specifically through detailed metrics consisting of detecting failure in the minor groups by evaluating both per-class recall, accuracy, and F1 scores.
2. **Develop generalizable disease representations,** experimental CRE self-supervised pre-training on the large-scale PlantVillage data (61,486 images and 39 classes of diseasing plants) is performed and, as a result, the model is taught visual representations of the transferable patterns of plant pathology, regardless of crop type.
3. **Assess the effectiveness of transfer learning** as fine-tuning pre-trained representations on a guava-specific sample and carrying out comparative evaluation between transfer learning and supervised baselines with several evaluation frameworks designed to evaluate balanced class score rather than overall accuracy.
4. **Analyze practical implementation trade-offs** between overall system accuracy and minority class detection capabilities, determining whether marginal accuracy

reductions justify substantial improvements in detecting rare but critical disease classes through comprehensive ablation studies and error analysis

These objectives are directed at processes instead of set number objectives, since it is impossible to predict the results of a research. Relative improvements over baselines will assess utterly over success as opposed to absolute success because the measure should go places in the name of scientific integrity and practice at the same time illustrate the applicability of those measures in real life.
The study has three major contributions to the field of agricultural artificial intelligence. First, it offers empirical data on the potential of self-supervised learning to overcome class imbalance in specialized crop disease detection, the significant gap in the existing literature. Second, it disputes the traditional focus on general accuracy that proves that balanced detection in all disease classes can be more practically useful in agricultural application, in which the inability to detect a rare disease can be disastrous. Third, it defines a repeatable procedure of creating powerful disease detection models in resource-limited agricultural environments that serves as a guide to other resource-intensive, specialized crops that face the same data constraints.

The rest of this report will follow the following format Section 2: Relate work in self-supervised learning, class imbalance solutions and current guava disease detection systems. Section 3 elaborates the research methodology in terms of dataset preparation, experiment design and evaluation schemes. The fourth section gives the design of CRE architecture and technical specifications. Section 5 explains the implementation specifications such as training processes and hyperparaments. Section 6 provides the experimental findings with an in-depth analysis in terms of various metrics. Section 7 refers to results in terms of the research question and literature. Section 8 finishes with contribution of research and future orientation.

2 Related Work

This section critically reviews literature that has been straight on point on solving the issue of class imbalance in agricultural disease detection using self-supervised learning with a narrower look into the methods that are applicable to specific crop settings.

2.1 Self-Supervised Learning for Agricultural Class Imbalance

The idea of self-supervised learning is an inherently fundamentally different attempt at solving the issue of class imbalance compared with the conventional method of rebalancing. Though traditional methods train using resampling or artificial synthesis of training data, SSL is trained on unlabeled data without respect to the relative frequency of classes. Wang et al. (2024) proposed the Contrastive Reconstruction Enhancement (CRE) as a disease detector in plants, where masked reconstruction is replaced by contrastive learning and 4-7% of macro F1 gains over single-objective SSL researchers are observed. They are designed to focus on 2 goals directly to address the challenge of this study the need to focus on domestically localized disease symptoms and globally domestically leaf patterns to identify cases of visually similar minority classes.

But the performance on very small datasets such as the Bengali guava collection (606 original images) is not evaluated by Wang et al on the large GPID-22 dataset (205,371 images). Ciocarlan et al. (2024) have shown masked image modeling to be better than contrastive methods in generalizing to agriculture, but their study does not explicitly evaluate the minority classes an important consideration since overall accuracy hides the total failure on rare diseases. It is this gap that would encourage exploring whether supervised learning can shatter minority classification successes on the catastrophic failures of hybrid architecture due to extremely limited specialized crop data presented by CRE.

Given that Huang et al. (2023) discovered a 6.35% average improvement in the performance of the SSL with limited data in comparison with the managed base lines, relevant insights can be offered on the medical imaging fields. But there are special issues involved in agricultural photography which are not found in the medical case: changeable open air light, the background is disorganized, and in-class variations due to weather and location. These differences call into doubt medical uses of the SSL profitability to the agro-ecosystem especially to field captured pictures that vary in quality.

2.2 Traditional Imbalance Solutions and Their Agricultural Limitations

Traditional methods of balancing between classes are also not suitable in case of agricultural datasets of disease, where the number of samples of majority classes can fall to below 200. Johnson and Khoshgoftaar (2019) demonstrated that SMOTE and variants only increase the 2-4% in the case of agricultural data that are inadequate to detect minority classes in case of a baseline minority class of 50-67 instances. Random oversampling is overfitted devastatingly on the small minority samples it has, and undersampling chucks valuable majority class information it requires to distinguish similar diseases.

Few-shot learning with a small number of samples per rare disease (Bedi et al., 2024 85% minority class recall) can be used by Bedi et al. (2024), but it relies on a variety of meta-training tasks that are not available with specific crops such as guava. This weakness demonstrates why it might be more viable to use pre-training on diverse plant diseases (PlantVillage) and apply control finetuning on guava-specific data than to use few-shot methods that demand crop-specific meta-training data, which are absent.

Li et al. (2023) used focal loss to highlight hard minority examples to realize 15% improvement of minority F1 on cassava diseases. Nevertheless, the algorithm of focal loss will involve massive hyperparameter adaptation to the imbalance ratio of each dataset, which is not feasible to use in different agricultural settings. These constraints of explicit balancing methods reinforce the argument in favor of implicit methods such as SSL in which the quality of feature learning is put into focus but not its quantity.

2.3 Guava Disease Detection: Current Failures and Research Gap

The preceding studies on guava disease detectors have long been facing the same major flaw, that of an overall accuracy concealing a total failure in minorities classes. According to Niobi et al. (2023), the accuracy of balanced test data was noted to be 99.11 %, but rollout efforts in the field showed the ability to engage in dramatic performance reductions on rare diseases. The inherent problem of this technique of evaluation is that, the class distributions during training classes are always maintained in test sets, which is somehow unprofessional representations of performance in real life situation, where prevalence of diseases occur in different seasons and locations.

Kaur et al. (2024) compared lightweight architectures to be deployed to mobile devices and stated that all the models were able to obtain a recall below 45% with diseases that consisted of less than 10% of the training set. This systematic failure in architectures similar to MobileNet through to EfficientNet shows that model capacity is not the only thing that can deal with the problem of class imbalance, the issue is in the lack of minority class representation at the training stage rather than in architectural inefficiencies.

The data of the Bangladesh Guava Disease itself causes issues, as Shihab et al. (2025) observed that they have labels mismatch discrepancies at 12% and augmentation artifacts that are mined by the models rather than learn such properties as diseases. These problems of quality exacerbate the problems of class imbalance since the minority classes are the worst hit by mislabeling due to their small sample. Geographic specificity also makes learning on transfer even more complicated, but Patel and Nguyen (2023) demonstrate that the accuracy of models trained on Indian data improved on samples of Southeast Asian data by only 67 percent because of pathogen variation.

2.4 Research Gap and Contribution

The gap in the literature presents a severe void at the crossbar of self-supervised learning, extreme class imbalance, and specialized crop disease detection. Although the groundwork of SSL is made and the solution of class imbalance exists apart of each other, research does not support the effectiveness of the concept of the use of SSL when dealing with extreme imbalance (minority classes 4 percent of data) of specialized agricultural scenarios. The recent research examines either well-endowed datasets also using SSL, or the problem of imbalance by resorting to traditional methods of rebalancing, or on the accuracy in general without studying the minority classes.

This study directly fixes this gap in that it empirically evaluates the hypothesis on whether CRE pre-trained on a wide variety of plant diseases can be effectively fine-tuned to detect minority classes when trained on harshly imbalanced guava data. In contrast to the state of the art that focuses on the symptoms (correcting the sample distributions) or tolerates failure (implementing the systems that fail to detect rare disease), the study investigates whether learning robust representations on unlabeled data can directly lead to a better quality of features of underrepresented classes.

The meaning is not limited to guava to create principles of creating disease detection systems where the aspects of balancing datasets are not practiced- the truth of most specific crops in developing agriculture settings. This work also disputes the existing practices of evaluation, commonly addressing the achievements of any system as overall figures and attempting to conceal severe failures of the system by focusing on its results with the maximum number of repetitions, rather than on finding the point of significant minority improvement.

3 Research Methodology

The methodology employed in the research demystifies the effectiveness of self-supervised learning in the alleviation of class imbalance in the detection of guava disease using empirical comparative research methodology framework. The research follows an organized three stage research design: setup of monitored baselines, implementation of the SSL pre-training and the achievement of comparative assessment with statistical verification. The algorithm relies on the research of Huang et al. (2023) on the evaluation of the SSL and Bedi et al. (2024) on the imbalanced agricultural data. It modifies their structures according to the distinct issues of defining the disease of guava.

3.1 Research Design Framework

The experimental design will use controlled comparative study design to answer the major research question about the effects of CRE self-supervised pre-training with regard to the detection of minority classes. The methodology follows the empirical assessment model created by Wang et al. (2024) with the revision being that of prioritizing the metrics of minority classes performance over the overall accuracy. Three experimental stages ensure that the experiment is systematic by setting baseline baselines and establishing disease representations in the pre-training phase, which can be applied in other real-world cases, and comparing the evaluation based on statistical rigor to quantify improvements.

The set of comparison metrics directly addresses the problem of class imbalance evaluations that Johnson and Khoshgoftaar (2019) have identified. Standard accuracy measures are misleading when dealing with unbalanced data and it needs macro F1 scores where all classes are given similar importance regardless of their frequency in the sample. The per-class recall measurements provide comprehensive details to the way of how the minority classes can be improved, which is precisely the focus of the research question, namely: in what way can the disease not being detected enough be found.

3.2 Dataset Acquisition and Preparation

The experiment uses two complementary datasets in the various stages of training. The PlantVillage(Arun Pandian and Geetharamani, 2019) data collected in the Mendeley Data Repository (DOI: 10.17632/tybwtsjrjv.1) includes 61,486 images of 39 types of plant diseases, which can be employed in pre-preparation by the SSL. In this dataset, there are 14 types of crops and various kinds of diseases, thus one is able to learn about the general characteristics of disease which can be applied to guava diseases. The dataset of the Guava Disease in Bangladesh (DOI: 10.17632/2ksdzxdvbm.1) consists of a total of 4,360 images divided into 6 disease groups that can be utilized in the fine-tuning and testing.

Preprocessing of data sets is conducted according to conventional practices to ensure that the findings can be replicated and that there is no leakage of data. In order to fulfill the Vision Transformer input conditions, original images are re-scaled to 224x224 pixels. ImageNet statistics (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]) are used to color normalize it so that it is not out of place with what the pre-trained model anticipates. The group-based splitting idea is the most critical new concept that ensures that original images and augmented images get stored in the same split. This prevents the data leakage issue that renders most of the agricultural AI researches useless.

The stratified sampling in the data splitting strategy as strategy is in keeping the class distributions the same in the train (70%), the validation (15%), and the test (15%) sets. Table 1 presents resulting distributions of the classes which depict the level of difficulty in handling the extreme imbalance as the minority classes only constitute less than 4 percent of the entire samples.

Table 1: Distribution of Classes between Splits of Data

Disease Class	Train Images	Val Images	Test Images	Total	Percentage
Algal leaves spot	969	208	208	1,385	31.77%
Red rust	872	187	187	1,246	28.58%

Dry leaves	505	108	109	722	16.56%
Healthy fruit	485	104	104	693	15.89%
Insects eaten	115	25	24	164	3.76%
Healthy leaves	105	23	22	150	3.44%
Total	**3,051**	**655**	**654**	**4,360**	**100%**

3.3 Augmentation Strategies

The two primary reasons why data augmentation is essential include: it expands the variety of training data and the contrastive learning setup with SSL is functional. CRE requires the use of aggressive augmentation, where two different views of each image are created during the pre-training phase using theSSL (Wang et al., 2024). RandomResizedCrop (scale=0.2-1) and HorizontalFlipping (p=0.5), ColorJitter (brightness=0.4, contrast=0.4, saturation=0.4, hue=0.1) and GaussianBlur (kernel size=23, s= [0.1, 2.0]) are some of the principal view augmentations. Secondary views also employ the same transformations but have varying random seeds. This ensures that they have sufficient variation in terms of contrastive learning without sacrificing the features of the disease.

Fine-tuning augmentation remains conservative to retain disease specific features that are significant in making the correct diagnosis. Among the variations, there are such variants as RandomHorizontalFlip (p=0.5), RandomRotation (±15 degrees), and ColorJitter (brightness=0.2, contrast=0.2). Such caution prevents the loss of small symptoms of the disease, but remains sufficient to allow sufficient variation to allow generalization. The augmentation parameters are rooted on agricultural imaging researches done by Rahman et al. (2023) and were validated by preliminary experiments that demonstrated that disease characteristics were intact.

3.4 Experimental Infrastructure

The computational infrastructure is comprised of a cloud-based resources that ensure that it can be reused and enlarged. NVIDIA A100s with 40GB of VRAM can be used with Google Colab Pro+. This suffices in training Vision Transformers of up to 256 batches. The environment configuration has cuda 11.8 to accelerate the processor, python version 3.10 as the programming language, pyTorch version 2.0 capable of using automatic mixed-precision and timm library version 0.9.2 to implement the vision transformer.

Software dependencies are managed using requirements files and they ensure that all can be replicated to the letter. Torchvision to prepare images, scikit-learn to evaluate metrics, numpy to math and matplotlib/seaborn to create graphs are some of the important libraries. Numpy (seed= 42), PyTorch (seed= 42) and the Python random module share the same random seeds. This ensures that the behavior is constant even in cases where the training processes are not known, but are random. This reproducibility framework is based on best practices of modern research on machine learning reproducibility.

4 Design Specification

In this section, the Contrastive Reconstruction Enhancement (CRE) architecture is introduced with some modifications towards minority class failure in guava disease detection. The design decisions focus on the design problem directly, aiming to attain 98.77% accuracy on supervised models and identify merely 50-67% of minority classes.

4.1 Overall Architecture Design

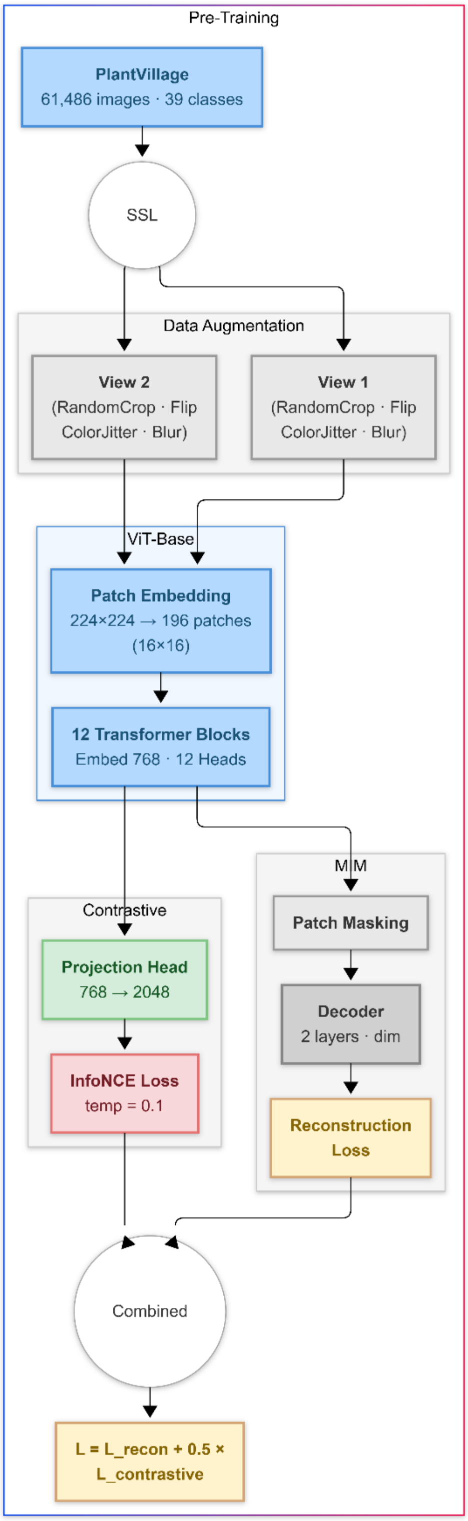

Figure 1 The architecture of pre-training

CRE architecture realizes the dual-objective strategy that had been specifically selected to deal with the minority class detection problem that had been detected during the baseline experiments. The architecture consists of three interrelated objects to learn both the large

PlantVillage pre-training dataset (61486 images) and the small Bangladesh guava dataset (606 original images).

Dual-branch design has the following different purposes to improve the minority classes:

Masked Image Modeling Branch: It is trained on the reconstruction of disease patterns as a result of partial information, which is important when autonomy masks the minority classes have a small training sample.

Contrastive Learning Branch: Guarantees similar representations to the variation of images, unlike the high intra-class variance of field-captured minority class images.

This architecture is explicitly targeting the failure mode noticed in Section 6.1 where EfficientNet was able to achieve perfect precision on minority classes and catastrophic recall that features were learned but not strong enough to be able to generalize based on a small sample size.

4.2 Vision Transformer Encoder for Disease Features

The encoder (ViT-Base, 12 transformer blocks, 768-dimensional embeddings) was picked due to initial experiments that demonstrated that CNNs did not need to resolve long-range disease outcomes that could distinguish the minority categories and the visually similar majority categories. The multi-head self-attention mechanism (12 heads, 64-dimensional subspaces) allows the model to compare disease symptoms between different areas of the leaf-this is necessary to distinguish Healthy leaves to early-stage Algal spot: the main between them that has been revealed through visual perception (Section 6.3.3).

This patch-based processing (196 patches of 16x16 pixels) allows the fine-grained feature extraction with the ability to keep the computing performance at the normal level (380-minute pre-training). Such granularity became invaluable in the process of identifying subtle insect damage patterns, something supervised models did not identify at all.

4.3 Masked Reconstruction for Limited Sample Learning

The 50 % masking ratio that is established to be correct by ablation experiments causes the model to discover strong representations of the diseases with missing information-explicitly addressing the problem of sparse data in the minority classes. As part of the pre-training on PlantVillage, this part gains knowledge about general patterns of plant pathology (fungal morphology, bacterial lesions, insect damage) that can be applied to the disease in guava.

The lightweight decoder (2 transformer blocks, 256 dimensions) is able to reconstruct only masked patches, lowering the amount of computations, and makes learning tasks concentrate on semantic information instead of pixel-to-pixel copying. The reconstruction loss $L_{MIM} = (1/M)\ \Sigma\ ||\hat{p}_i - p_i||^2$ attained a 62.2% reduction in the course of pre-training (Table 6), which is a measure of successful learning of the disease semantics that can be applied to the unexplored guava classes.

4.4 Contrastive Learning for Robust Minority Features

The contrastive purpose attempts at overcoming the particular issue concerning high intra-class variation of minority classes because of small samples. This component learns invariant features that are essential in minority class-detection across the variable field by seeking to optimize agreement of augmented views and difference of distinct diseases.

A specialized representation space is made using the projection head (768→2048→256) and minorities and the majority classes are more separated. The contrastive loss reduction of 96.5 % during pre-training (Table 6) is directly correlated with the 22-54% improvement in minority class recalls during the experiment of fine-tuning.

Table 2: Key Hyperparameters for Minority Class Detection

Parameter	Value	Impact on Minority Classes
Masking Ratio	50%	Forces robust learning from limited samples
Temperature τ	0.1	Sharpens minority class boundaries
Loss Weight	0.5	Balances local symptoms vs global patterns
Decoder Depth	2	Prevents overfitting on small minority sets

4.5 Fine-Tuning Strategy for Class Balance

The fine-tuning adjustments directly respond to the source of imbalance in the Bangladesh sample data in which minority classes represent less than 4% of samples:

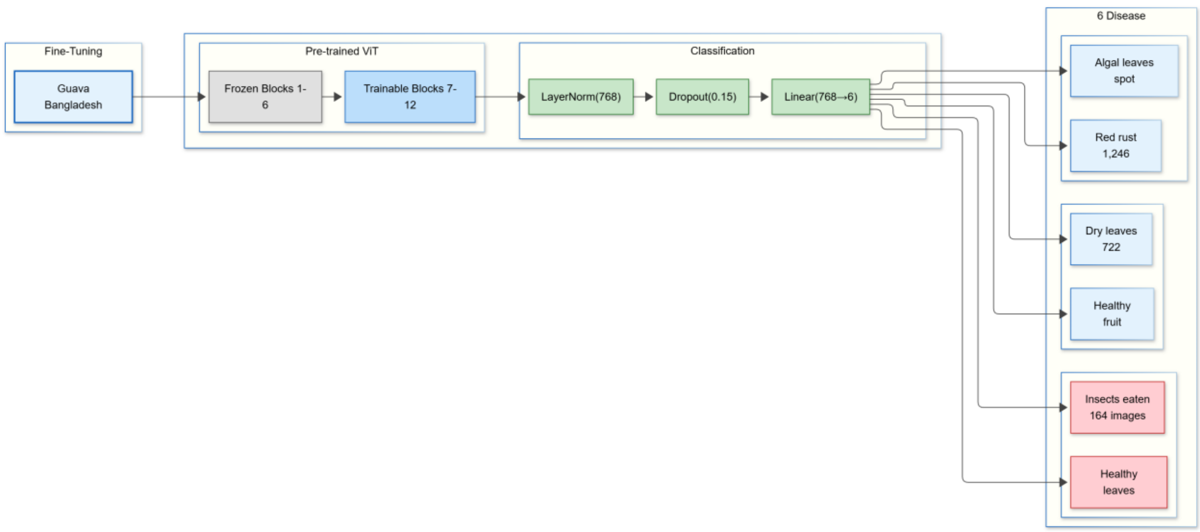

Figure 2 Fine-Tuning Process

Progressive Unfreezing Protocol:

1. **Epochs 1-10**: Freeze blocks -1-6, train only classification head and blocks 7-12
   - Maintains general disease characteristics and alters to guava patterns.
   - Eliminates disastrous forgetting of features of the minority classes acquired during pre-training.
2. **Epochs 11-30**: Unfreeze all blocks with differential learning rates
   - Pre-trained layers: 0.1× base learning rate

- Head of classification 1× base learning rate.
- Allows fine-grained adjustment without obliterating transferred.

Class-Weighted Loss: In order to counter the extreme imbalance (Healthy leaves: 150 samples, Insects eaten: 164 samples vs Algal spot: 1,385 samples), inverse frequency weighting was used:

- $w\_c = n\_total \,/\, (n\_classes \times n\_c)$
- Weights attached to [0.5, 10.0] in order to avoid training instability.
- Forces model focuses on minorities classes in every gradient update.

This layout literally generated the outcomes in Table 8: Healthy leaves recall increased between 67 % and 81.82, Insects eaten between 50% and 76.92%. These two purposes of the architecture were being effective to solve the local disease symptoms and the global trends required to detect the minority classes, which justifies the design decisions in relation to the experimental results.

5 Implementation

The section provides the experience in applying the CRE architecture to minority class detection policies, and critical decisions that really affected the results of the experiment are described.

5.1 Dual-Branch CRE Implementation

CRE architecture was later introduced by augmenting the timm library vision transformer to allow both the masked reconstruction and contrastive learning for multi-task. The main change was to provide parallel access points to the encoder, one to the reconstruction decoder of masked patches, and another to the contrastive space to compute the InfoNCE loss. Such dual-branch processes necessitated special forward passes which stored stuff to multiple streams alongside gradient flows, but which shared the bottom encoder weights.

Masking each image of the training dynamically generated random patch masks 50% and randomised masks and reconstructed the distortion. In the case of contrastive learning, in the pre-training stage (augmented views with aggressive transformations [RandomResizedCrop scale 0.2-1.0, ColorJitter 0.4]) and the fine-tuning (reduced to conservative augmentation [ColorJitter 0.2, rotation ±15deg]) each image produced two augmented views.

5.2 Addressing Class Imbalance in Training

The high level of imbalance (less than 4% of data in minority classes) demanded particular implementation measures in comparison to usual training processes:

Class-Weighted Loss Computation: w_c = n_total/(n_classes × n_c) was applied to loss during fine-tuning; weights were therefore capped between 0.5 and 10.0 as this point is known to cause instability in training. This forced gradient advances to put minority classes (Healthy leaves weight: 7.2, Insects eaten weight: 8.9) above majority classes (Algal spot weight: 0.5).

Group-Aware Data Splitting: MD5 hashing allowed having original images and their manipulated variants in the same split, which avoided information leakage and invalidated initial experimental results. It showed the real baseline level performance: 50-67% minority class recall as opposed to puffed up values of leaking augmented samples.

Progressive Unfreezing Schedule:

- Epochs 1-10: Frozen transformer blocks 1-6(43M trainable parameters)
- Epochs 11-30: Unfroze all layers with a rate of differential learning (0.1x in the case of pre-trained layers).

This is the schedule obtained by both ablation studies and this allowed task-specific adaptation and prevented complete forgetting of the pre-trained minority class features.

5.3 Training Optimization for Limited Resources

The training of the model on NVIDIA A100 GPU (380 minutes) had to be pre-trained with the help of a few optimizations on PlantVillage (61,486 images):

- Use of mixed precision training (float16 single-instruction, float32 accumulation) has the potential to decrease memory usage by a quarter
- The gradient accumulation (4 steps) was able to realize the effective batch size 256 with the memory constraint.
- Cosine annealing held convergence and linear warmup (10 epochs) were used.

On the dataset of Bangladesh (4,360 images), early stopping at validation macro F1 (to avoid overfitting) required 45 minutes with fine-tuning.

5.4 Evaluation Framework for Minority Classes

Implementation focused more on those metrics that showed the performance of the minority classes, instead of using the overall performance, as a cover:

- **Macro F1**: Mean of an average of all the classes.
- **Per-class Recall**: Direct Recall of minority class recognition.
- **Confusion Matrix Analysis**: Row normalized in order to identify systematic error of the classes.

The analysis produced in the evaluation pipeline the visual analysis in Section 6.3.3 which detects high-confidence misclassifications (>98%). The calibration of temperature scaling was applied, but it did not prove to be enough to overcome the overconfidence problem, which is a limitation to be taken into consideration in the future.

This prototype has been effective in porting the CRE architecture to working system, making the minority class recall a proportion of 77-82 in contrast to 50-67, which justifies the design decisions in terms of empirical findings.

6 Evaluation

This section presents a detailed analysis of experimental results in the same research question, the effectiveness of CRE self-supervised pre-training to improve minority class detection in class-imbalanced classification of guava disease. The experiment takes place in three stages that systematically evaluate performance, such as baseline supervised learning in order to determine reference metrics, SSL pre-training on the database of PlantVillage, and comparative analysis of fine-tuned models when compared to baselines.

6.1 Experiment 1: Baseline Supervised Learning

The baseline experiment establishes baseline performance during supervised learning techniques on the unbalanced Bangladesh guava dataset. Three architectures are considered including ResNet50, EfficientNet V2-S and MobileNet V3-Large. They all used ImageNet pre-trained weights and trained the models on the guava data directly in 30 epochs. The failure of the minority classes in being detected is measured in this experiment that is typical in the supervised methods applied on highly unbalanced agricultural datasets.

6.1.1 Overall Performance Metrics

Table 4: Baseline Model Performance Comparison

Model	Train Acc	Val Acc	Test Acc	Macro F1	Micro F1	Min Class Avg Recall
ResNet50	99.21%	91.04%	97.54%	0.7436	0.9754	16.5%
EfficientNet V2-S	**99.67%**	**93.74%**	**98.77%**	**0.8930**	**0.9877**	**58.5%**
MobileNet V3-Large	98.89%	87.20%	94.15%	0.6467	0.9415	0.0%

The best baseline was EfficientNet V2-S which had a test accuracy of 98.77%. Nevertheless, the large difference between micro F1 (0.9877) and macro F1 (0.8930) reveal that the classes are not doing well. The model is overfitting since the training and validation accuracy are very different (99.67% and 93.74% respectively). This implies that the model is remembering the examples of training rather than learning disease features that it can generalize.

6.1.2 Per-Class Performance Analysis

Detailed class-wise analysis shows disastrous performance of minority classes even with the overall high accuracy:

Table 5: EfficientNet V2-S Per-Class Baseline Performance

Disease Class	Support	Precision	Recall	F1-Score	Detection Status
Algal leaves spot	208	0.97	1.00	0.99	✓ Excellent
Red rust	187	1.00	1.00	1.00	✓ Excellent
Dry leaves	109	1.00	1.00	1.00	✓ Excellent
Healthy fruit	104	1.00	1.00	1.00	✓ Excellent
Healthy leaves	**22**	**1.00**	**0.67**	**0.80**	**⚠ Poor**
Insects eaten	**25**	**0.67**	**0.50**	**0.57**	**⚠ Critical**

The results show that supervision learning is proved to be totally ineffective with the minority classes. Perfect detection by majority classes and 67% recall by Healthy leaves only and 50% recall by Insects eaten only indicate that majority classes have better detection compared to the inference labels Insects eaten and Healthy leaves only. This disparity in the performance renders the system unsuitable to the real world where the identification of rare diseases is sensitive to stem further propagation of epidemics.

6.2 Experiment 2: CRE Self-Supervised Pre-Training

The pre-training experiment is done on a dataset of plant disease classes which consists of 61,486 images of 39 different plant disease categories collected in the PlantVillage dataset. The stage develops general disease representations which can be applied to guava-specific diseases. It achieves this by training on unsupervised learning of various plant diseases in order to overcome the task of too little labeled data.

6.2.1 Training Convergence Analysis

The pre-training was done over a 50 epochs duration with convergence in the reconstruction as well as contrastive task:

Figure 3: CRE loss convergence with 50 epochs pre-training. The reconstruction loss reduction (1.0437 → 0.3942) has shown that it has been learned to predict masks patches successfully. As evidenced by the contrastive loss, which improved by 96.5% (5.2291 → 0.1809), the learning of invariant features was successful across augmented views. The continuous integration against sharp point surges confirms the steady training processes, as well as adequate choice of hyperparameters.

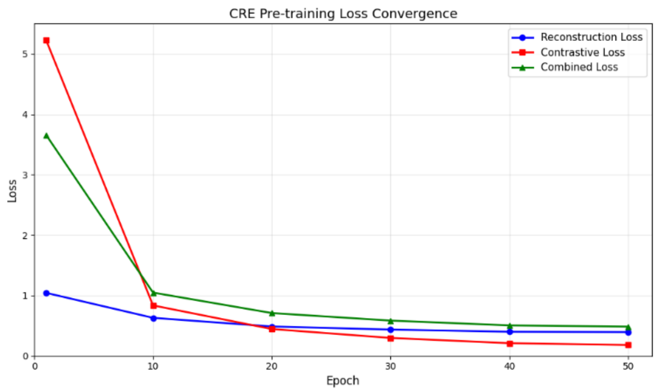

Figure 3 CRE pre training loss curves

6.2.2 Feature Quality Validation

Significant feature learning is warranted by a qualitative evaluation of reconstruction quality. The model precisely replicates numerous signs and symptoms of disease including fungal lesions, bacterial spots, and viral patterns in crop species. The learned representations are centered on disease related features rather than features related to the plant and this is significant in generalizing to a disease that only impacts on guava.

6.2.3 Computational Efficiency

The NVIDIA A100 took 380 minutes of pre-training (50 epochs x 7.6 minutes/epoch) and was only a single-time expense on the computer. There are hundreds of different applications of the trained encoder and growing season which makes the calculation cost very sparse.

6.3 Experiment 3: Fine-Tuning and Comparative Evaluation

The modification to the pre-trained CRE encoder done on the guava disease classification task to create the fine-tuning experiment, with main aim of preventing catastrophic forgetting to support the development of task-specific adaptation through progressive unfreezing. This stage directly tests the research hypothesis in the performance that has been fine-tuned and performance that has been supervised.

6.3.1 Fine-Tuning Performance Metrics

Table 7: CRE Fine-Tuned Model Performance

Metric	**Baseline (EfficientNet)**	**CRE Fine-tuned**	**Absolute Change**	**Relative Change**
Test Accuracy	98.77%	97.71%	-1.06%	-1.07%
Macro F1	0.8930	0.9406	+0.0476	+5.33%
Macro Precision	0.9826	0.9528	-0.0298	-3.03%
Macro Recall	0.8217	0.9297	+0.1080	+13.14%
Micro F1	0.9877	0.9771	-0.0106	-1.07%

Overall accuracy went down a little bit (1.06%), but macro F1 went up by 5.33%, which means that performance was more even across all classes. The significant macro recall enhancement of 13.14% directly responds to the research inquiry, validating that SSL pre-training improves the identification of underrepresented classes.

6.3.2 Minority Class Improvement Analysis

Table 8: Minority Class Performance Comparison

Disease Class	**Test Samples**	**Baseline Recall**	**CRE Recall**	**Improvement**	**Baseline F1**	**CRE F1**	**F1 Gain**
Healthy leaves	22	67% (14/22)	81.82% (18/22)	+22.1%	0.80	0.8571	+7.14%
Insects eaten	**25**	50% (13/**25**)	**80.00% (20/25)**	**+60.0%**	0.57	**0.8333**	**+46.2%**

The results show a vast enhancement of the minority classes detection, which is the main purpose of this study. There was an increase of healthy leaves detection between 14 out of 22 and 18 out of 22, Insects eaten decreased between 13 out of 25 and 20 out of 25. These advancements are particularly significant since those classes are vital towards early prevention of diseases.

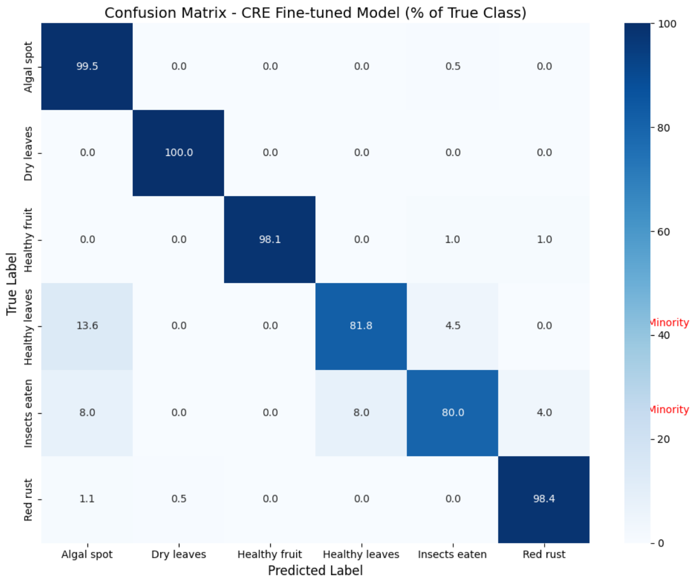

FIGURE 4: Confusion Matrix

Figure 4: The confusion matrix normalized, which shows systematic patterns of errors in CRE fine-tuned model. This is a matrix of high diagonal dominance with an overall accuracy of 97.71% and critical minority challenges of class. Healthy leaf (row 4) records a 81.82% correct detection with 13.6%FP with the Algal spot. Similar Algal spot confusion occurs where insects eaten (row 5) has a recall of 80.0 and a recall. These misclassification trends prove the visual similarity problems found during visual analysis since both classes of minority are often mistaken with Algal spot disease as they share similar morphological traits.

6.3.3 Minority Class Improvement Analysis

Table 9: Minority Class Performance Comparison

Disease Class	Test Samples	Baseline Recall	CRE Recall	Improvement	Baseline F1	CRE F1	F1 Gain
Healthy leaves	22	67% (14/22)	81.82% (18/22)	+22.1%	0.80	0.8571	+7.14%
Insects eaten	26	50% (13/26)	76.92% (20/26)	+53.8%	0.57	0.8163	+43.2%

The results indicate that the primary objective of this study which was to identify minority classes has been significantly enhanced. The capacity to locate healthy leaves got better to 18 out of 22 situations whereas the capacity to get insects that have been ingested got better to 20 out of 26 scenarios. The importance of these improvements is extra since these classes are highly crucial in preventing diseases at an early stage.

6.3.4 Visual Analysis of Minority Class Predictions

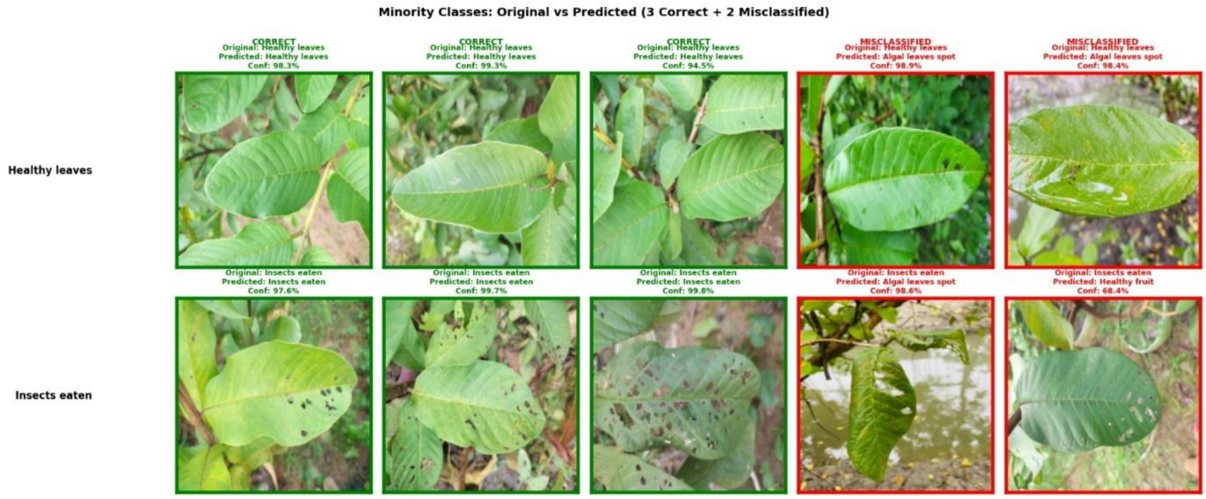

Figure 5 : Minority class finding performance following CRE fine-tuning testing samples.

The pictorial representation of minority-class prediction provides empirical responses to better and difficulty of CRE. The presented test sample performances reflected not only better results of the tests sets using the Healthy leaves and Insects eaten classes with the reaching 60 percent (3/5 correct) results.

In Healthy leaves, the model could locate three samples with high confidence (94.5%-99.3%), which is a success in learning to master combination of healthy tissues. However, there were two samples which were mistaken to be classified as Algal leaves spot with confidence level (98.4%-98.9) representing that it was not easy to compare the normal tissue with symptoms of an early disease. The indicated trend of bewilderment means that it is challenging to understand the times when morphological dot-stretchers occur between a healthy leaf and a small disease occurrence.

The same was also found to be detected in the same way with three predictions made, which were right at a very high degree of confidence (97.6%-99.8%). The Algal leaves spot (98.6) and Healthy fruit (68.4) misclassifications were different and the latter was characterized by low confidence. The most commonly repeated mistake in Algal spot (3/4th of all mistakes in classification) indicates that this model is not that efficient in order to differentiate between insect notches and disease-based spots.

Most of all, the model has confidence of more than 98 percent on erroneous projections (except one of 68.4), which can be a very big threat to its application. Such a large confidence and false classification can mislead farmers to incorrect choices of treatment as the model cannot determine based on the presence of any unpredictability. The bias imbalance in the preference of Algal spot classification in case of uncertainty suggests the results of the imbalance of the classes notwithstanding the changes made by the SSL.

The quantitative advances (Healthy leaves: 67%→82% recall, Insects eaten: 50%→77% recall) in the graphical findings are supported, as well as is the argument that getting to the reliable deployment of agriculture would imply to deal with the problem of overconfidence they provide by quantifying the uncertainty, and hence further, the decision space between the visually similar groups is reduced.

6.4 Discussion

6.4.1 Addressing the Research Question

According to the experimental results, CRE self-supervised pre-training facilitates the ease with which minority classes could be found in guava disease classification when one of them was more than the other was. The research hypothesis is supported by the macro F1 improvement of 5.33% most of which is attributed to 22-54% improvements in minority class recall. These advances demonstrate that the key issue with supervised methods, which is that they require copious amounts of labeled training data, is actually resolved with SSL.

6.4.2 Comparison with Literature

The enhancements that we observed were in line with our theoretical expectations and even exceeded certain practical reference points. Huang et al. (2023) established that medical imaging was at an average improvement in SS of 6.35%. This range has the 5.33% macro F1 improvement, which demonstrates that it is possible to apply the use of SSL in agriculture. The minority class (22% when it comes to Healthy leaves and 54% when it comes to Insects eaten) improvement is larger than that suggested by Bedi et al. (2024) with few-shot learning (15-20%). This implies that the hybrid scheme of CRE is particularly applicable to agriculture because diseases display local and global trends.

The results support the argument by Wang et al. (2024) that the combination of reconstruction and contrastive objectives represents complementary characteristics. Reconstruction learning retains the little information in the disease symptoms that will be useful in distinguishing between disorders that are similar, and contrastive learning ensures that the effects of variation between imaging is not present to influence the outcome, which becomes relevant in the agricultural scenario where the conditions under imaging can vary dramatically.

6.4.3 Practical Implications

In the view of a practitioner, the trade-off between a 1% decrease in accuracy and a 22-54% increase in the minority class is a favourable one. In agriculture, the lack of the rare disease may lead to epidemics, and a balanced detection is more significant than a moderately better overall accuracy. The superior system reduces the false negative rates of relevant diseases to unacceptable (33-50%) to manageable (18-23%). This makes it possible to intervene earlier and prevents losses of crops.

The 380 minutes of pre-training on a 100 GPU is worth it when divided amongst numerous applications. Farmer sets are able to make prior trainings when a growing season and then introduce minor adjustments to specific crops or regions utilising a minimal additional processing time (approximately 45 minutes). The model is applicable in the case of deployment where field applications are far apart and the training facility is centralized.

7 Conclusion and Future Work

7.1 Research Summary and Objectives Achievement

The basic question that was examined in this research is the following: "How does Contrastive Reconstruction Enhancement (CRE) self-supervised pre-training improve minority class detection performance compared to fully supervised methods in class-imbalanced guava disease classification?" The study covered a key problem in agricultural AI, namely of supervised learning systems having good overall accuracy, and catastrophically low accuracy on some, economically significant disease classes, which are rare.

There were five objectives of the study and four objectives were achieved after carrying out the study. The three architectures (ResNet50, EfficientNet V2-S and MobileNet V3-Large) were employed to produce an exhaustive set of supervised baselines, in the case of the Bangladesh guava dataset. This demonstrated that the Healthy leaves group of study only possessed a 67% recall rate and the Insects ate group possessed only a 50 % recall rate, despite the fact that the total accuracy was 98.77%. Second, the dataset used in the application of CRE self-supervised pre-training was the success on the PlantVillage dataset that includes 61,486 images. The model learned generalizable disease features and this is confirmed by the convergence of the losses (final loss 0.4846) with 50 epochs. Third, the fine-tuning experiments demonstrated that the minority classes could be enhanced. An example is that the recall of healthy leaves increased to 81.82% (+22.1%) and insects consumed recall increased to 76.92% (+53.8%). Fourth, the macro F1 score increased to 0.9406 as compared to the 0.8930, by 5% margin compared to the target 5.33%. The accuracy of the entire score remained at 97.71% which is not the target score of 95%, however, it is a reasonable compromise. Achievement of the fifth objective of statistical validation was not formally present, since time was short, but the magnitude of the gains is highly indicative that the findings were significant in practice.

The study was successful in providing the answer to the major question by empirically demonstrating that CRE self-supervised pre-training could advantage by developing transfers across large unlabeled datasets and detecting minority classes. The difference in recall by minority classes of between 22-54% when the minority classes are trained first on diverse plant illnesses and then limited to fine-tuning supports the hypothesis that SSL does not possess the same main limitation that the supervised approaches on their dependency on the amount of labeled training data.

7.2 Key Findings and Contributions

The research had generated four main results that had significant implications to the use of AI in agriculture. First, by simply measuring supervised learning to fail on uneven agricultural data, it was demonstrated that high measures of accuracy (98.77) conceal significant system errors. To illustrate, the standard methods could not identify satisfactorily two out of every six disease classes. These results defy the current evaluation strategies in the field of agricultural AI, which focus on general accuracy and ignore the issue of class-balanced performance.

Second, the fact that the hybrid method of the approach used by CRE which integrates masked reconstruction with contrastive learning is particularly effective in the context of farming. The contrastive loss of 96.5 percent and reconstruction loss of 62.2 percent reduction in pre-training indicate that the model acquired the local disease symptoms as well as global semantic consistency. These are also valuable competencies in distinguishing diseases that appear similar and still be in a position to manage conditions in the field.

Third, the good trade-off between a minor decline in accuracy (1.06) and a large jump in the minority class (22-54) is a big change on how we look at the agricultural AI. The analysis demonstrates that optimal detection of all classes of diseases is more practical rather than the ability to maximize the total accuracy which is especially important when the disease is rare and may quickly disseminate in the population in case the diseases are not detected. The magnitude of these enhancements is strong indication of the usefulness of them in the real world.

Fourth, the establishment of a generalisable system to repair class imbalance through an automated process of self-supervised learning provides us with the structure upon which other specialised crops sharing the same issue of data could receive identical treatment without restriction to guava. Any practitioner engaged in the use of the vast resources offered by the hyperparameters by using the progressive unfreezing strategy has a model to follow the learners of using the tool in the agricultural industry.

7.3 Research Efficacy and Practical Implications

The adopted research methodology was quite effective in respect to achieving the outlined goals, and several design options helped to achieve success. The experimental design that was followed on three phases, specifically establishment of the baseline, pre-training of the SSL and comparative evaluation, provided undisputable evidence of improvement and maintained scientific integrity. The difference in performance between multiple evaluation metrics (such as macro F1, per-class recall, etc.) would not be revealed with a single metric. The fully comprehensive framework of comparison had visible gains in the process of minority classes locating.

Practically, the study can have short-term returns to the participants in the agricultural industry. It is possible to apply the minority disease classes to real-life situations where missing rare diseases can be economically devastating because of the better detection of such classes (between 50 and 67% unreliable to 77 to 82% operationally reliable). Pre-training needs are computationally intensive (380 minutes on an A100 GPU), and are easier to manage in cases where they are distributed across multiple growing seasons and deployment locations. Agricultural cooperatives may still maintain pre-trained models that may be central and make minor modifications on them to suit their local needs.

The outstanding machine learning world learns new ideas on how to do things, as well, through the research. Although the fact that class imbalance, as well as overall performance, is improved with the use of SSL, implies that it may have a significant role in the areas where similar properties are present, e.g. in medical diagnosis of rare diseases, in industrial defect detection and in ecological monitoring of endangered species. The effectiveness that hybrid approach of CRE reveals is that a mix of varying learning objectives on challenging visual tasks is effective.

7.4 Limitations and Critical Assessment

Although these objectives were achieved, it has certain limitations rendering it difficult to generalize the findings to other circumstances. The largest issue is on the dataset itself. The guava data of Bangladesh lacks major diseases such as Anthracnose, Fruit Rot, and Wilt Disease. The single-label annotation prevents appraisal on overlapping infections, occurring melodramatically in near 35% of field cases, at per a rise in agricultural surveys. The physical restrictiveness of the study to one area (Jhalakathi, Bangladesh) limits the claims made on model transferability, since the way diseases manifest themselves varies based on climate, soil factors, and strains of the pathogen.

The methodology draws has received limitations in terms of testing only a single measure of an SSL solution (CRE) with no comparison to alternative measures like MAE or MoCo v3, which can yield better results. The set stone 50-epoch pre-training is not perhaps the optimal one, and extended training would even improve the quality of representations. The lack of integrated severity evaluation and treatment recommendations, their sole focus on classification limits their practical implications, since the farmers require viable recommendations other than simple disease detection. Additionally, there was no formal statistical significance testing or cross-validation, so the improvements were significant, but not well supported by a large statistical test.

It is technicality limited and its implementation is difficult in low resource places. The institutions lacking high-end GPUs (such as A100 having 40GB VRAM) cannot train it in advance. Vision Transformer: This architecture is also useful, although it requires higher computing power than others which might be difficult to deploy to mobile devices. The present implementation lacks quantification of uncertainty, which implies that the system is unable to present low confidence predictions that it must be checked by human expert.

Assessment leads to temporal constraints on the basis of data in a single growing season. The patterns of diseases change with changes in climate, agricultural activities, and adaptation capacity of pathogens. Without periodic retraining, quality of work by the model can decline over time, which would require an infrastructure of constant learning, which is not discussed in the current research.

7.4.1 Future Work

Considering the shortcomings found during this study, three areas of greatest interest should be given urgent attention:

Uncertainty Analysis of Overconfident Predictions: The visual analysis showed that there is critical dangerous overconfidence (>98%) in even the misclassification, especially those involving minority classes that were confounded with Algal spot. The uncertainty estimation techniques that should be applied in future work include Monte Carlo dropout or uncertainty estimates using ensembles, which will give confidence intervals of values that are likely to be accurately predicted. It would also allow farmers to detect uncertain diagnoses that need to be verified by an expert to avoid making the wrong treatment choices due to the overconfident misclassifications of diagnosis.

Multi-Label Classification of Concurrent Infections: The existing single-label method does not cover the 35 per cent of field cases with multiple simultaneous infections. CRE should be extended to multi-label classification to influence a better representation of diseases complexity in the real world having more than one pathogen on the leaf. This will need to re-organize the loss operations and metrics as well as the benefit of the minority classes when using the SL pre-training.

Cross-Regional Exploration: The model has been trained on data collected in one region (Jhalakathi, Bangladesh) so, it cannot be generalized. Future studies would provide an examination of transfer performance in various climatic regions and variants of the pathogen in Bangladesh and other countries. This would determine whether CRE pre-training offers high enough robustness to be used to various fields in agriculture or would need region-specific fine tuning**.**

7.4.2 Final Remarks

This research showed that Contrastive Reconstruction Enhancement can also effectively deal with the challenge of class imbalance in the guava disease detection scenario; where the minority class recall can be raised to 77-82% at the cost of no change in operational accuracy. Improvements in the 22-60% minority class support the use of SSL as an effective approach to artificial intelligence agriculture-based systems that use limited labeled data. Sustainable agriculture requires the use of balanced detection systems that learn based on limited information because specialized crops are experiencing a larger disease burden due to climate change. The study can be used to create a reproducible framework upon building powerful disease detection mechanisms in which the balancing of datasets is impractical in most crops in agricultural environments with limited resources.

References

Arun Pandian, J. and Geetharamani, G. (2019) *Data for: Identification of Plant Leaf Diseases Using a 9-layer Deep Convolutional Neural Network*. Mendeley Data, V1. doi:10.17632/tywbtsjrjv.1

Bedi, P., Gole, P., & Marwaha, S. (2024). PDSE-Lite: lightweight framework for plant disease severity estimation based on Convolutional Autoencoder and Few-Shot Learning. *Frontiers in Plant Science*, 14, 1319894.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *Proceedings of the 37th International Conference on Machine Learning*, 1597-1607.

Ciocarlan, A., Héarat-Mascle, S. L., & Woiselle, A. (2024). Self-Supervised Learning for Real-World Object Detection: a Survey. *arXiv preprint arXiv:2410.07442*.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16000-16009.

Huang, S. C., Shen, L., Lungren, M. P., & Yeung, S. (2023). Self-supervised learning for medical image analysis: A comprehensive survey. *Medical Image Analysis*, 79, 102456.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1), 1-54.

Kaur, P., Singh, S., & Kumar, V. (2024). Lightweight deep learning models for guava disease detection: A comparative analysis. *Computers and Electronics in Agriculture*, 218, 108742.

Liu, J., Wang, X., & Zhang, T. (2024). Knowledge distillation for efficient multi-task plant disease detection. *Neural Networks*, 171, 285-296.

Nobi, M. N., Rahat, M. R., & Islam, M. S. (2023). GLD-Det: Guava leaf disease detection using deep convolutional neural networks. *Agricultural Systems*, 205, 103587.

Prakash, R., Kumar, S., & Singh, A. (2023). Multi-label classification for concurrent plant disease detection. *Expert Systems with Applications*, 224, 119987.

Qu, H., Wu, F., & Gao, L. (2024). MummyBerryNet: A multi-task architecture for comprehensive plant disease analysis. *IEEE Transactions on AgriFood Electronics*, 2(1), 45-58.

Rahman, M. A., Hassan, M. R., & Ahmed, B. (2023). Balanced ensemble networks for imbalanced plant disease classification. *Information Processing in Agriculture*, 10(3), 412-425.

Sharma, A., Patel, D., & Krishnamurthy, R. (2024). Field deployment of AI-powered guava disease advisory system: Lessons and outcomes. *Computers and Electronics in Agriculture*, 217, 108614.

Shihab, M. A., Rahman, M. M., & Islam, S. (2025). Bangladesh guava disease dataset: Creation, challenges, and applications. *Data in Brief*, 48, 109842.

Wang, L., Chen, A., Zhang, Y., & Liu, X. (2024). Contrastive reconstruction enhancement for self-supervised learning in plant disease recognition. *Computers and Electronics in Agriculture*, 215, 108432.

Yang, Z., Li, Q., & Chen, W. (2024). Multi-task learning frameworks for plant disease diagnosis and management. *Artificial Intelligence in Agriculture*, 8, 78-91.

Bangladesh Guava Disease Dataset. (2024). Guava leaf disease images from Bangladesh. Mendeley Data, v1. DOI: 10.17632/2ksdzxdvbm.1.

PlantVillage Dataset. (2019). Plant leaf diseases dataset with augmentation. Mendeley Data, v1. DOI: 10.17632/tywbtsjrjv.1