AgriEngineering

Article

Addressing Class Imbalance in Rice Disease Diagnosis with Contrastive Dissimilarity

Fabio Pignelli [1,*], Yandre M. G. Costa [1,2], Loris Nanni [3] and Lucas O. Teixeira [2]

[1] Graduate Program in Computer Science, State University of Maringá, Maringá 87020900, Brazil; yandre@din.uem.br
[2] Department of Informatics, State University of Maringá, Maringá 87020900, Brazil
[3] Department of Information Engineering, University of Padua, 35131 Padua, Italy; loris.nanni@unipd.it
* Correspondence: fabio.pignelli@gmail.com

Abstract

Rice disease recognition is an important task in agricultural diagnostics, as it directly affects crop health and yield. The challenge intensifies due to the visual similarity between symptoms of different diseases and pests, as well as the inherent class imbalance within real-world datasets, where some conditions are more represented than others. In this paper, we explore the application of contrastive dissimilarity for rice disease identification using the Paddy Doctor dataset. Our approach specifically targets the challenges of uneven class distribution, aiming to improve recognition accuracy in data-scarce scenarios common in precision agriculture. Through an extensive set of experiments, we demonstrate that our approach generates robust and discriminative representations, even under conditions of imbalanced and/or underrepresented data, reaching 98.4% on accuracy.

Keywords: rice disease; disease identification; agricultural diagnosis

1. Introduction

Rice is one of the most important crops worldwide, providing the primary source of calories for more than half of the global population. However, its production is constantly threatened by a wide range of diseases caused by fungi, bacteria, and viruses, such as rice blast, sheath blight, and bacterial leaf blight [1]. These diseases can lead to severe yield losses, significantly affecting food security and farmers' income, particularly in developing countries. Early and accurate diagnosis of paddy diseases is, therefore, a key step in implementing timely control measures and ensuring sustainable rice production.

Traditional methods of plant disease diagnosis rely on manual inspection by trained specialists, which is time-consuming, subjective, and impractical at large scales [2]. In recent years, advances in artificial intelligence (AI) and computer vision have enabled the automatic detection and classification of plant diseases from leaf images, providing an alternative to conventional practices [3]. Convolutional Neural Networks (CNNs) and related deep learning architectures have achieved remarkable results in image-based recognition tasks [1,4]. Although their performance depends on the availability of large, balanced, and well-annotated datasets, conditions are rarely met in real-world agricultural environments.

In paddy disease diagnosis, the problem is intensified by class imbalance and limited data availability, as certain diseases occur more frequently or are easier to capture than others. This imbalance may bias the learning process toward dominant classes, leading

Check for updates

Academic Editors: Yee Siang Gan, Sze-Teng Liong and Ray E. Sheriff

Received: 8 November 2025
Revised: 25 January 2026
Accepted: 29 January 2026
Published: 2 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

to poor generalization for rare diseases. Recent work highlights that class imbalance is a pervasive issue in precision-agriculture machine learning and can make overall accuracy misleading because models optimized for total error tend to favor majority classes. In particular, Miftahushudur et al. [3] survey imbalance-handling strategies across agricultural applications and emphasize using imbalance-aware metrics, including F1-score, and techniques that explicitly target minority-class performance, including resampling and synthetic data generation.

Recent research in computer vision and pattern recognition has highlighted the importance of methods that can capture both similarities and dissimilarities among samples in a discriminative feature space. In this context, metric learning approaches have demonstrated strong potential, as they map samples into representations where semantic relationships are encoded through the distances between them. The recently proposed contrastive dissimilarity framework [5] has demonstrated promising results in various domains, particularly in scenarios with imbalanced and limited data [6–8]. By integrating representation learning with a dissimilarity-based contrastive metric, this approach offers robustness to unbalanced class distributions and improved generalization from scarce samples.

This work evaluates the applicability of contrastive dissimilarity [5], a method designed to address data imbalance and limited sample availability. This approach integrates a dissimilarity-based representation with contrastive learning into a unified framework, using representation learning to extract meaningful features and metric learning to estimate a task-specific dissimilarity function. Using the Paddy Doctor dataset [9], which provides annotated images of diseased and pest-affected rice leaves, we evaluate the capacity of this method to discriminate between visually similar conditions and to mitigate the adverse effects of dataset imbalance.

In summary, this work makes three contributions: (i) we propose contrastive dissimilarity for rice disease diagnosis, especially for imbalanced datasets, as it is in real scenarios; (ii) we provide an empirical analysis showing that our approach can perform well even for diseases that have few samples in the original dataset, and (iii) we demonstrate consistent improvements with different ablations to verify the robustness of our approach across multiple configuration.

The remainder of this work is organized as follows: Section 2 presents some related works. Section 3 brings some details about the materials and methods, and also about the experiments, including the datasets and training protocol. Section 5 describes the results obtained, and following, we point out some concluding remarks achieved with this work in Section 6.

2. Related Works

Miftahushudur et al. [3] provide an overview of methods for imbalanced agricultural datasets, organizing approaches into algorithm-level, data-level, and hybrid strategies, with an emphasis on resampling pipelines (e.g., over/under-sampling, SMOTE-style variants) and their trade-offs (e.g., overfitting, boundary overlap, computational cost). The survey also reviews the emerging use of deep generative models for synthetic augmentation in high-dimensional data, and discusses open challenges to field deployment, such as noisy and incomplete data, difficulty of early-stage disease separation, and limited availability of standardized public benchmarks for reproducible comparisons.

Zhang et al. [10] introduced an ensemble framework that embeds a salient-position attention mechanism into different lightweight backbones (YOLO, EfficientNet, MobileNet, ShuffleNet) and fuses them. The approach improves single-model performance and reaches 98.33% accuracy for rice-leaf disease identification in complex field backgrounds, highlighting gains in robustness and generalization under natural conditions.

Mookkandi et al. [11] propose a lightweight vision transformer with 814.7 k learnable parameters and 85 layers, designed for classifying crop diseases in paddy and wheat. The architecture consists of a convolutional block attention module, squeeze and excitation (SE), and depth-wise convolution, followed by a ConvNeXt module. The proposed model was tested using a paddy dataset (with 7857 images and eight classes) and a wheat dataset (with 5000 images and five classes). In this work, the author report an accuracy of 98.47% for the paddy dataset, and 92.8% for the wheat dataset.

Using a custom dataset with 5932 diseased rice leaf images and 1500 healthy images, Petchiammal and Murugan [12] evaluated deep Convolutional Neural Network (CNN) and nine transfer learning models (VGG19, VGG16, DenseNet121, MobileNetV2, DenseNet169, DenseNet201, InceptionV3, ResNet152V2, and NASNetMobile) using TensorFlow. Each model's performance was assessed to find the most effective classification system, covering four disease categories and one non-disease category. The research aimed for shorter training times, higher accuracy, and easier retraining, with DenseNet121 achieving the highest classification accuracy of 97.6% on the paddy leaf image dataset.

Padhi et al. [13] present the usage of EfficientNet B4, a deep learning architecture trained on the publicly available Paddy Doctor dataset. They used data augmentation to reach 19,131 labeled images for training and 4785 images for testing. The model aims to classify paddy leaf samples into nine disease categories or into normal specimens. EfficientNet B4, known for its structured scaling technique and Swish activation function, was trained using pretrained weights from ImageNet and optimized with the Adam optimizer, achieving an accuracy of 96.91%.

Also using the Paddy Doctor dataset, Thanuboddi and Nelakuditi [14] proposed the usage of Self-Supervised Deep Hierarchical Reconstruction (SSDHR), and Long Short-Term Memory (LSTM), which perform early disease detection based on the spatial and temporal data, respectively. The SSDHR network uses multi-branch convolution kernels to extract distinct discriminative characteristics rather than conventional leaf-based indicators. It incorporates spatial, and temporal-based attention mechanisms, Symmetric Fusion Attention, to improve feature selection and XGBoost. The proposed approach was evaluated using the Paddy Doctor dataset with 16,225 sample images from 13 classes, and experiments show that the framework achieves a 99.25% accuracy rate.

Bera et al. [15] introduce an attention-based CNN framework for plant-disease recognition by focusing on the diseased regions via region-wise feature aggregation and attention weighting. They benchmark on several public datasets, including Paddy Doctor, and report accuracy across datasets when using standard CNN backbones (e.g., DenseNet/MobileNet) with ImageNet initialization. Their best reported accuracy result using Paddy Doctor is 99.65%.

Furthermore, in the work of Petchiammal et al. [9], the Paddy Doctor dataset was proposed, and also benchmarking results were proposed using the same dataset variant as we used in our work. The benchmarking results, using different models, are presented in Table 1.

Table 1. Related results in Paddy Doctor dataset, reported by [9].

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
CNN	89.22	88.84	88.81	88.84
MobileNet	92.63	92.42	92.39	92.42
VGG16	93.49	93.19	93.2	93.19
Xception	96.61	96.58	96.57	96.58
Resnet34	97.52	97.50	97.50	97.50

3. Materials and Methods

This section details the methodological framework employed in this study, which integrates dissimilarity-based representation learning, metric learning, and the contrastive dissimilarity (CD) approach. The methodology aims to enhance classification robustness in scenarios characterized by class imbalance and limited data availability. Initially, we describe the theoretical foundations of dissimilarity representation and its connection to metric learning, followed by the implementation of the contrastive dissimilarity model. This combined approach enables the model to learn task-specific distance functions that enhance feature discrimination across visually similar disease classes, thereby addressing one of the primary challenges in rice disease diagnosis.

3.1. Dissimilarity

Dissimilarity-based representations emphasize the differences between samples rather than their absolute features, providing an alternative perspective for feature encoding [16]. This approach is efficient when structural variations are highly discriminative, as observed in applications such as texture classification and handwriting recognition.

Two principal formulations are typically employed: dissimilarity space and dissimilarity vector [17]. In the dissimilarity space approach, each sample is described by its dissimilarities to a set of representative prototypes, thereby embedding the data into a new feature space suitable for conventional classifiers. On the other hand, the dissimilarity vector approach computes pairwise feature differences between samples and prototypes, transforming the task into a binary classification problem that evaluates whether two samples belong to the same class.

3.2. Metric Learning

Metric learning is a field dedicated to developing data-driven distance functions that better reflect task-specific relationships between samples [18]. Conventional distance measures, such as Euclidean, Manhattan, or cosine, are often selected based on prior domain knowledge; however, they may fail to capture the complex, nonlinear structures present in real-world data. For instance, in image classification or pattern recognition, an effective distance function must account for both local and global features to accurately assess similarity [19].

To overcome the limitations of predefined metrics, metric learning methods aim to automatically learn an optimal distance function from supervised or semi-supervised data. This process adapts the metric to the underlying data distribution, enhancing the discriminative capability of models and improving overall performance. Metric learning has been widely applied in numerous machine learning tasks, including classification, clustering, and retrieval, demonstrating substantial improvements across different domains.

3.3. Contrastive Dissimilarity

Contrastive dissimilarity is an approach designed to overcome the limitations of imbalanced and limited data scenarios in pattern recognition [5]. Unlike conventional techniques that rely on static similarity measures, such as Euclidean or cosine distance, this method employs a task-specific metric learning process within the contrastive learning framework, enabling the calculation of dissimilarity functions better adapted to the specific scenario it addresses.

Furthermore, contrastive dissimilarity has been successfully applied to different application domains, such as Music Classification [6] and Writer Identification [7].

Presented by Teixeira et al. [5], the proposed schema has three sequential stages: metric learning, Siamese training, and dissimilarity representation. In the first stage, we define the

model architecture before training, with a focus on the projection head, which serves as the core component responsible for learning the task-specific metric.

The second stage performs siamese network training. For tabular datasets, each sample (row) is paired with another instance from the same class, and this operation is repeated to produce as many pairs as determined by the batch size. For each pair, the element-wise absolute difference is computed and passed through the projection head, which outputs a dissimilarity score used to calculate the contrastive loss. During training, all possible pairwise combinations are generated internally, eliminating the need for manually annotated negative pairs.

The third stage focuses on generating the dissimilarity representation. Here, the trained projection head replaces conventional, fixed-distance metrics by estimating data-driven dissimilarity values. Each observation in the dataset is compared against a subset of representative prototypes, forming the dissimilarity-based training and testing matrices. These prototypes serve as reference points that capture the diversity and structure of the data distribution. Furthermore, dissimilarity can be categorized into the following two main types:

- Dissimilarity space, where each sample is represented as a vector of dissimilarities against the selected prototypes.
- Dissimilarity vector, which condenses each sample–prototype relation into a compact vector.

4. Experimental Setup

In this section, we describe the experimental setup designed to evaluate the effectiveness of the proposed contrastive dissimilarity framework in the rice disease diagnosis task. The experiments were conducted using the Paddy Doctor dataset, which offers a realistic and imbalanced representation of field conditions. We outline the dataset characteristics, data preparation, and model training protocols, as well as the evaluation procedures used to assess performance under varying levels of data scarcity. The objective is to verify whether the proposed approach maintains competitive accuracy and robustness across different training regimes and dataset configurations.

4.1. Paddy Doctor Dataset

The Paddy Doctor dataset (Available at [https://dx.doi.org/10.21227/hz4v-af08], accessed on 24 January 2026) was originally composed of visual and infrared images of paddy leaves obtained from real paddy fields in a village near the Tirunelveli district of Tamilnadu, India [9]. In this work, we explore the visual images aiming to perform automatic rice disease classification. The images were collected in real fields using high-resolution smartphone cameras between February and April 2021. After cleaning and annotation, the dataset contains 16,225 RGB images at a resolution of 1080 × 1440 pixels, organized into 13 classes: 12 disease categories and healthy leaves, as detailed in Table 2.

The class distribution in the dataset is imbalanced, ranging from 450 samples for Bacterial Panicle Blight (BPB) to 2405 samples for the Normal class. Excluding the Normal class, approximately half of the remaining classes contain fewer than 1000 samples, while two classes exceed 2000 samples. This uneven distribution indicates a disparity that may bias the diagnostic performance of machine learning models.

Furthermore, Figure 1 shows some sample images from the dataset.

The original authors of the Paddy Doctor dataset released four variants. In this work, we used the "Small and Split" variant, which consists of 16,225 images resized to 256 × 256 pixels and divided into training and test sets. The training set contains 12,980 images (80%), and the test set contains 3245 images (20%). Both sets were stratified according to class labels and paddy variety to ensure reproducibility.

Table 2. Samples distribution per class across the Paddy Doctor dataset.

Seq.	Paddy Disease Name	Count
1	Bacterial Leaf Blight (BLB)	648
2	Bacterial Leaf Streak (BLS)	505
3	Bacterial Panicle Blight (BPB)	450
4	Black Stem Borer (BSB)	506
5	Blast	2351
6	Brown Spot	1257
7	Downy Mildew	868
8	Hispa	2151
9	Leaf Roller	1095
10	Tungro	1951
11	White Stem Borer	1273
12	Yellow Stem Borer	765
13	Normal	2405
TOTAL		16,225

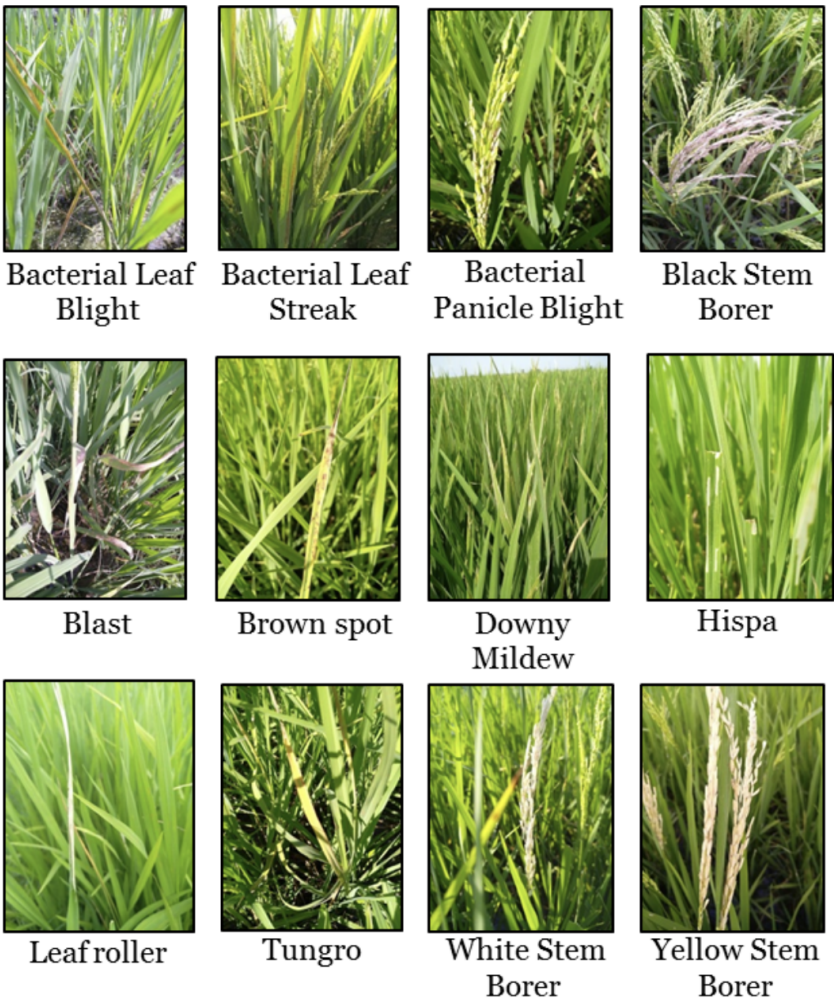

Figure 1. Sample disease images from the Paddy Doctor dataset. Adapted from ref. [9].

4.2. Training Protocol

The input images were from the Paddy Doctor dataset variant "Small and Split" (256 × 256). As illustrated in Figure 2 all dataset variants contain the same amount of images and classes. The only difference between them is image resolutions, which also impacted

the zip size, and the fact that the 'Small and Split" was already split into training and test sets. This variant was also used by Paddy Doctor's original author to run their benchmark.

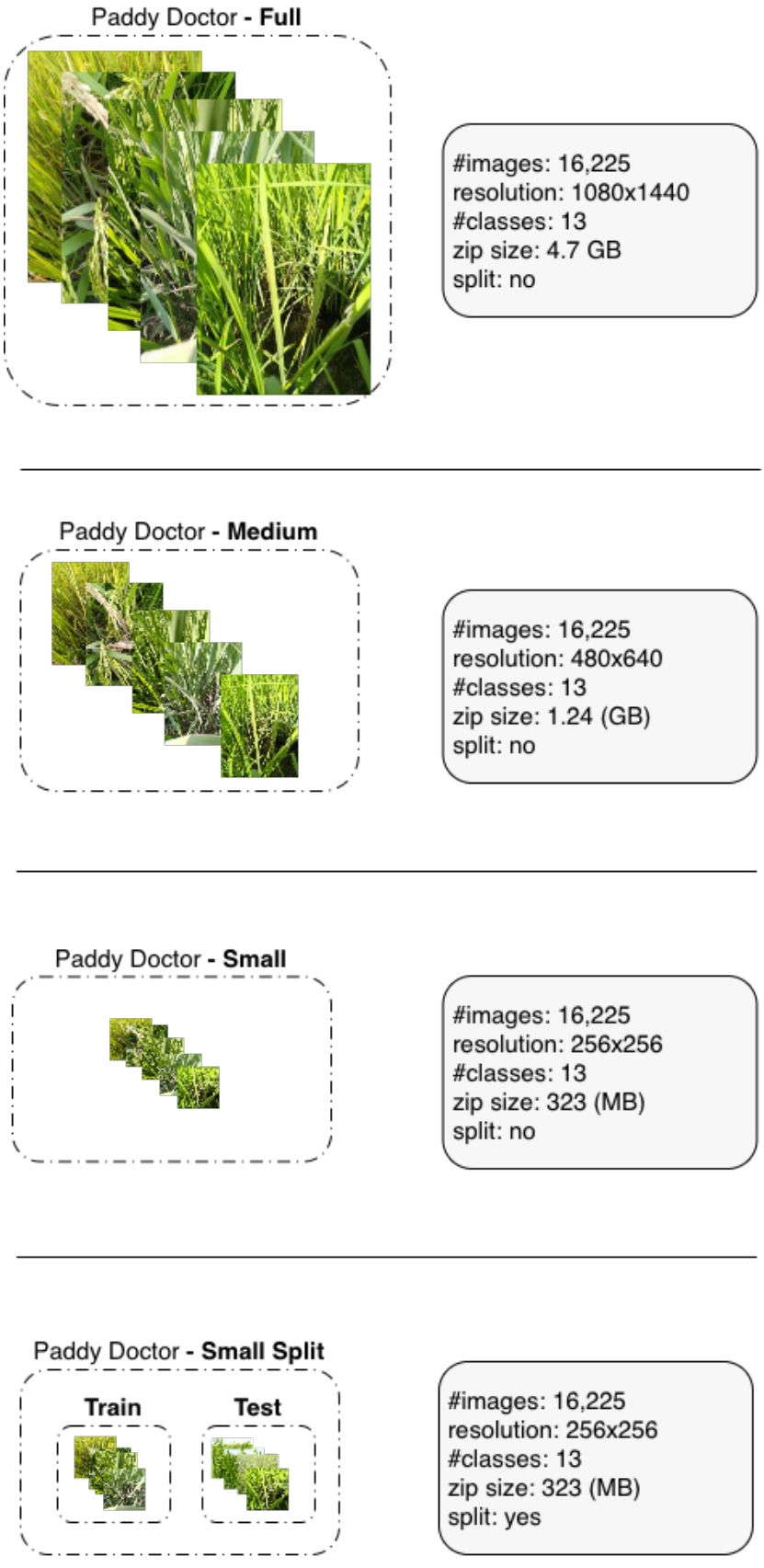

Figure 2. Paddy Doctor dataset official variants.

To train the proposed contrastive dissimilarity model, we adopted the same methodological foundations introduced by ref. [5], where the training is performed through a metric learning framework integrated into a contrastive loss function.

Algorithm 1 summarizes the pseudocode used in our approach, adapted from ref. [5]. The input consists of two sets, im' and im'', each containing data points. The batch size determines the number of matched pairs in each set, where elements at the same index belong to the same class (i.e., im'_1 matches im''_1, im'_2 matches im''_2, and so on). In Line 1, im' and im'' are concatenated into a single set x. Next, Lines 2–3 use tile and repeat to expand x into x' and x'': tile replicates the array along specified axes, while repeat duplicates each element a given number of times. This expansion is used to enumerate all possible pairings between elements originating from im' and im''. In Line 4, the method computes dissimilarities for the resulting pairs, and in Line 5 reshapes them into the matrix form required by the subsequent contrastive loss computation (Line 6).

Algorithm 1: Contrastive dissimilarity training

Input: image vector im' and im'', labels y, batch size n, model m
Output: updated model m

1: $x \leftarrow concatenate(im', im'')$
2: $x' \leftarrow tile(x, n * 2)$
3: $x'' \leftarrow repeat(x, n * 2)$
4: $D \leftarrow m.diss(x' - x'')$
5: $D \leftarrow reshape(D, (n * 2, -1))$
6: $L \leftarrow contrastive\ loss(D, y)$
7: $m.backward(L)$
8: **return** m

During training, the training set was augmented by creating sample pairs according to class membership, ensuring the presence of both positive (same-class) and negative (different-class) pairs. The dissimilarity function was learned through a projection head consisting of fully connected layers.

In addition, to simulate scenarios of limited data availability, we conducted experiments with progressively smaller portions of the training set: 100%, 50%, 20%, 10%, and 5% of the original training data. Notably, the testing set remained unchanged across all experiments, ensuring a fair and consistent comparison of results under varying training data conditions.

Figure 3 shows the classification setup we adopted in this work.

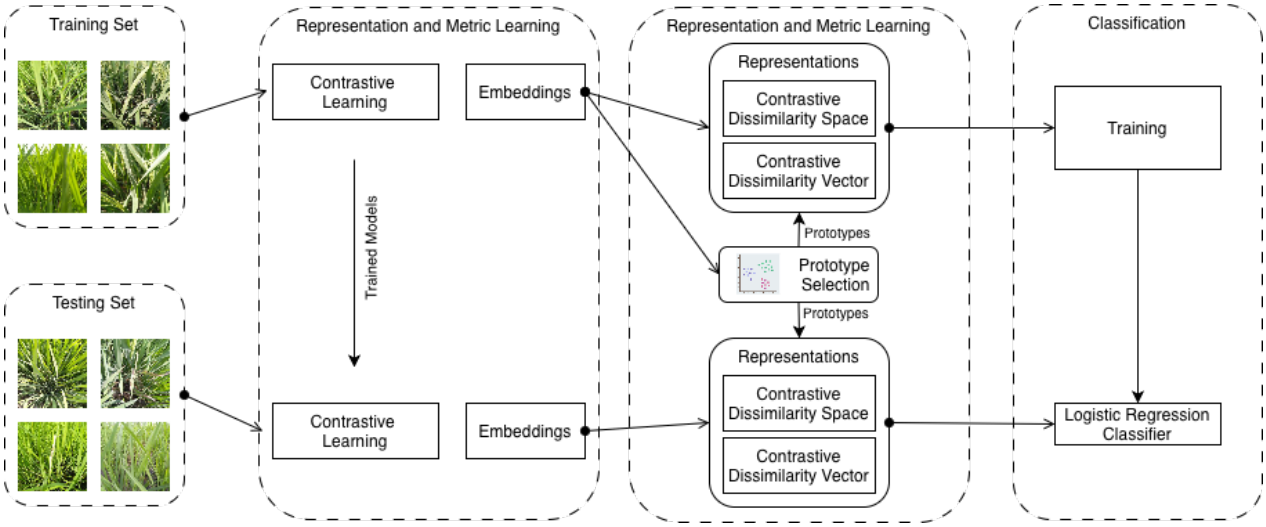

Figure 3. Classification setup for contrastive dissimilarity.

4.3. Evaluation Protocol

The evaluation of the proposed approach was conducted using the Paddy Doctor dataset, a benchmark dataset designed for rice disease classification. To ensure reproducibility and prevent bias, we followed the official "Small and Split" variant released by the dataset authors, which includes predefined training and testing partitions. Additionally, we considered the stratified variant provided with the dataset to ensure class representation.

To verify different levels of data scarcity, we repeated the evaluation across the exact progressively reduced training set sizes (100%, 50%, 20%, 10%, and 5%) while keeping the testing set unchanged in all scenarios.

For feature extraction, we employed EfficientNetV2 [20] as the primary convolutional backbone, while prototype selection was carried out using K-Means clustering. Additionally, we evaluated different embedding sizes for both the dissimilarity space and the dissimilarity vector.

For benchmarking, we compared our approach against a baseline Convolutional Neural Network (CNN), EfficientNetV2 [20], trained directly on the same dataset splits. This comparison enabled us to assess whether contrastive dissimilarity provided consistent improvements over a standard deep learning pipeline.

All representations were evaluated using Logistic Regression as a classifier. This interpretable yet straightforward classifier was chosen to keep the focus on assessing the viability of contrastive dissimilarity in the disease recognition context, without introducing further complexity.

The evaluation metric employed was classification accuracy, as it is the primary metric reported in the Paddy Doctor benchmark and widely adopted for image classification tasks.

5. Results and Discussion

This section presents the results obtained in our experiments, followed by a discussion that contextualizes our achievements with existing literature results.

Table 3 summarizes the results obtained under different training set sizes. As expected, the overall accuracy increased with the availability of more training data. At the most constrained scenario (5% of the training set), the baseline CNN achieved 80.6%, contrastive dissimilarity space (CDS, 79.6%), and contrastive dissimilarity vector (CDV, 79.8%). The difference was small (≈1%), demonstrating that the proposed approach remains competitive even under extreme scarcity conditions.

Table 3. Limited training set accuracy (%).

Training Size	[9]	CNN	CDS	CDV
5%	-	80.6	79.6	79.8
10%	-	86.3	86.4	86.6
20%	-	91.5	91.3	91.8
50%	-	96.1	96.3	96.1
100%	97.5	97.3	98.2	98.2

Table 4 shows that our proposed methods (CDS and CDV) consistently match or improve the baseline, with the largest gains observed in the low-data regime. At 10% training size, both variants improve precision (from 85.2 to 86.7 for CDS and 86.4 for CDV) and have higher F1-scores (84.7 to 85.2 and 85.3), indicating better overall balance between false positives and false negatives. The effect is even more evident at 20%, where CDV achieves the best performance across all metrics, reaching 93.0 precision, 90.9 recall, and 91.7 F1, surpassing the CNN baseline (92.0/90.3/91.1). As the training size increases to 50% and 100%, remains competitive when more data is available.

Table 4. Limited training set precision, recall and F1-Score (%).

Training Size	Precision				Recall				F1-Score			
	[9]	CNN	CDS	CDV	[9]	CNN	CDS	CDV	[9]	CNN	CDS	CDV
5%	-	81.0	82.4	81.7	-	76.7	75.7	76.3	-	78.4	78.2	78.4
10%	-	85.2	86.7	86.4	-	84.5	84.1	84.6	-	84.7	85.2	85.3
20%	-	92.0	92.5	93.0	-	90.3	90.3	90.9	-	91.1	91.2	91.7
50%	-	96.3	96.7	96.4	-	95.8	96.2	96.0	-	96.0	96.4	96.2
100%	97.5	97.4	98.3	98.3	97.5	97.3	98.3	98.3	97.5	97.4	98.3	98.3

With 10% and 20% of the training data, all models converged toward similar performance levels (≈86–91%), with CDV showing a small advantage, as seen in Figure 4. These results are consistent with the observations of [5], who reported that contrastive dissimilarity maintains stable accuracy across varying levels of data availability. At 50% training data, CDS surpassed the CNN baseline (96.3% vs. 96.1%), and at full training size (100%), both CDS and CDV outperformed the baseline CNN (98.2% vs. 97.3%), and also the benchmark reported by Petchiammal et al. [9] (98.2% vs. 97.5%).

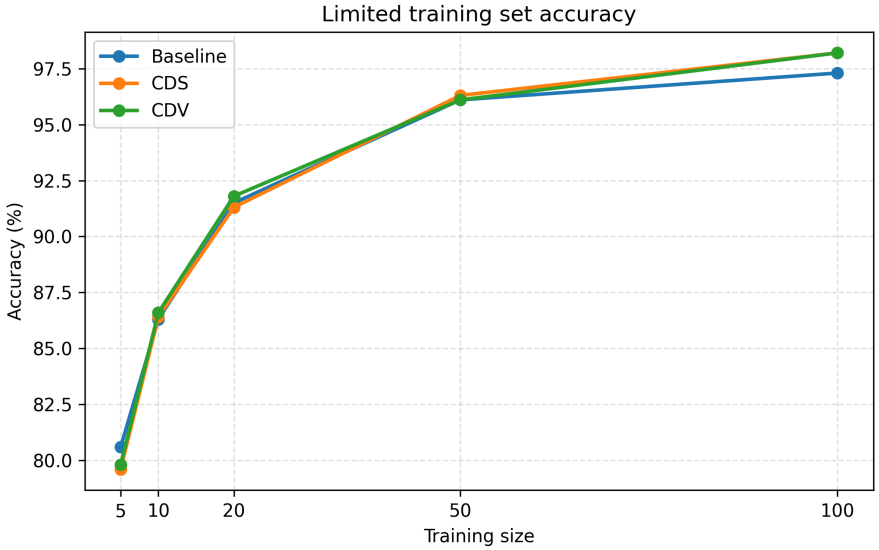

Figure 4. Limited training set comparison between accuracy on different training set sizes.

The results in the most constrained scenario is also shown in Figure 5, presenting the confusion matrix for training size with 5%.

Since both models predict on the same instances, we used McNemar's test, which is appropriate for paired classification outcomes. McNemar's test results were not significant at $\alpha = 0.05$, so we do not claim a significant overall difference. Still, the confusion matrix in Figure 5 shows the key effect: robust predictions for the minority class, which is the main outcome of our approach.

Table 5 presents the impact of different embedding dimensionalities on model performance. Accuracy remained consistently above 90% across all tested dimensions, with the best results achieved using compact embeddings (32-D for CDS at 98.4% and 128-D for CDV at 98.2%). Larger embeddings (256-D) did not show improvements. These findings are also consistent with those of [5], which stated that higher-dimensional projection heads do not necessarily enhance performance and suggested the use of compact embeddings instead.

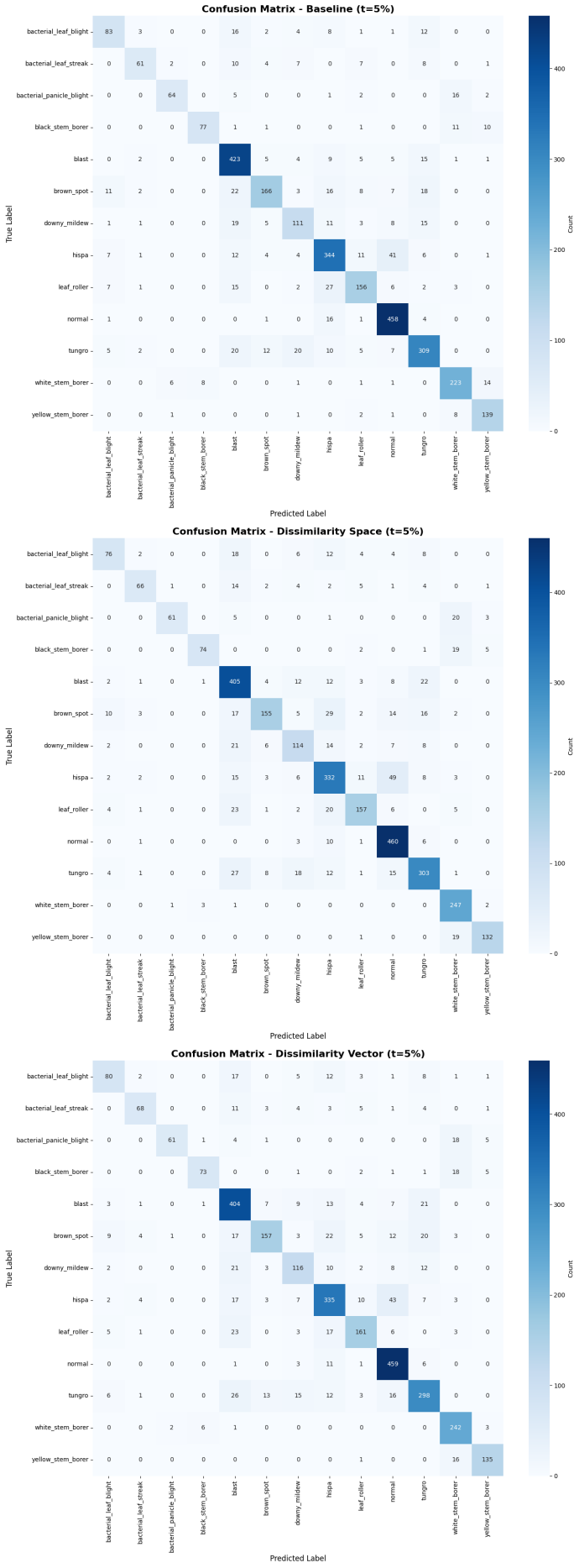

Figure 5. Confusion matrices for the training size with 5%.

Table 5. Embedding size impact on accuracy using CDS and CDV (%).

Embedding Size	CDS	CDV
32-D	98.4	98.0
64-D	98.2	97.8
128-D	98.2	98.2
256-D	98.0	97.9

To further investigate the robustness of our approach, we conducted an ablation study evaluating the influence of different data augmentation strategies on model performance. Three augmentation pipelines were tested: (i) no augmentation, consisting of random cropping; (ii) geometric transformations, vertical and horizontal flips, and rotations; and (iii) color-based augmentations, incorporating Gaussian blur and random brightness–contrast adjustments.

The results, summarized in Table 6, reveal that the overall accuracy of both CDS and CDV models remained stable across all configurations, with variations of less than 0.3 percentage points. This consistency indicates that the proposed framework is primarily invariant to the types of transformations.

Table 6. Data augmentation strategy impact on accuracy using CDS and CDV (%).

Strategy	CDS	CDV
No augmentation	98.2	98.4
Flips and rotations, only	98.2	98.1
Color augmentations, only	98.2	98.3

Table 7 shows that CDS and CDV are insensitive to the number of prototypes. Accuracy is already high with one prototype (98.1%) and quickly saturates, with only minor fluctuations thereafter. CDV is highest at two prototypes (98.4%), while CDS stays essentially the same, around 98.2%, suggesting that a small prototype set is sufficient.

Table 7. Prototype amount impact on accuracy using CDS and CDV (%).

# Prototypes	CDS	CDV
1	98.1	98.1
2	98.2	98.4
3	98.2	98.3
5	98.2	98.2
10	98.2	98.1
20	98.0	98.2

Table 8 shows that the choice of prototype selection algorithm has minimal impact on performance. All evaluated methods fall within a 0.2% range, with k-means giving the best overall results (98.2% for CDS and 98.3% for CDV). This indicates that our approach is not sensitive to how prototypes are selected, and simple clustering (k-means) is sufficient.

Table 8. Prototype selection algorithm impact on accuracy using CDS and CDV (%).

Algorithm	CDS	CDV
k-means	98.2	98.3
k-means++	98.2	98.2
spectral	98.1	98.1
hierarchical	98.2	98.1

In Table 9 we present a consolidation of results from the literature, our baseline and our approach. Across all configurations, our best-performing model achieves 98.4% accuracy, improving upon both the CNN baseline (98.4% vs. 97.3%) and the previously reported benchmark on the Paddy Doctor dataset (98.4% vs. 97.5%). We also evaluated a cost-sensitive (weighted) Support Vector Machine (wSVM), which introduces class-specific penalties by reweighting the hinge-loss so that minority-class errors are penalized more heavily than majority-class errors [21]. This approach is used for learning under class imbalance and asymmetric error costs; we employed the standard balanced weighting scheme, which sets weights according to the inverse class-frequency ratio. We trained the wSVMs with hyperparameters selected using grid search as follows: $kernel$ = ["rbf","linear"], $C = [0.0001, 0.001, 0.1, 1, 10, 100, 1000]$ and $\gamma = [0.0001, 0.001, 0.01, 0.1, 1]$. For each hyperparameter combination, model fitting and selection were performed using 5-fold cross-validation on the training set, and the final model was refit on the full training set using the best parameters identified during the search.

Table 9. Accuracy (%) of prior methods, our baseline, and the proposed approach.

Training Size	[9]	[14]	[15]	CNN + wSVM	CNN	CDS	CDV
5%	-	-	-	79.4	80.6	79.6	79.8
10%	-	-	-	86.1	86.3	86.4	86.6
20%	-	-	-	91.8	91.5	91.3	91.8
50%	-	-	-	96.0	96.1	96.3	96.1
100%	97.5	99.2	99.6	97.5	97.3	98.4	98.4

To ensure a fair comparison, the weighted SVM was trained on feature vectors extracted from the same baseline CNN used in our study. Under this protocol, the weighted SVM reaches 97.5% accuracy on the full dataset, falling short of our best result (98.4%).

In the constrained scenarios, the weighted SVM performs more competitively, surpassing the baseline CNN as well as the CDS and CDV variants. Notably, these gains stem from the control of the decision boundary via cost reweighting, whereas our method targets improvements at the representation level through a representation learning framework. In other words, the SVM's advantage in these settings reflects calibration of error trade-offs, while our approach seeks robustness through learned features that better capture minority-class structure under limited or constrained conditions.

Also, we contextualize our results against the published numbers available for the Paddy Doctor dataset. Direct comparison to the top-performing approach is limited, as prior work does not report performance under the same constrained evaluation protocol considered here. Nevertheless, under our stricter and more diagnostic setting, our method remains competitive and consistently improves over standard baselines, indicating that the proposed framework provides an effective and principled path for addressing class imbalance without relying solely on post hoc cost tuning.

Finally, we evaluated our approach by constructing a classifier ensemble combining CNN, weighted SVM, and our methods CDS and CDV using the sum rule for decision fusion [22]. Results are shown in Table 10, and we can observe that the ensemble consistently outperforms each individual classifier, indicating that the gains cannot be attributed to simple redundancy or averaging effects. In particular, while the CNN captures highly discriminative hierarchical representations and the weighted SVM emphasizes margin-based decisions with class imbalance awareness, the proposed approach contributes with information that is not fully exploited by either model alone. The improvement observed after fusion suggests that our method focuses on distinct characteristics of the data, leading to complementary patterns across classifiers. This diversity is a key requirement for

effective ensembles and provides empirical evidence that the proposed approach encodes novel and relevant information, rather than what is already learned by standard deep or kernel-based models.

Table 10. Accuracy (%) obtained using an ensemble of CNN, weighted SVM, and our proposed approaches.

Training Size	wSVM + CNN + CDS	wSVM + CNN + CDV	wSVM + CNN + CDS + CDV
5%	81.0	81.2	81.0
10%	86.9	86.9	87.0
20%	92.0	92.0	92.0
50%	96.4	96.3	96.5
100%	97.4	97.3	97.3

It is important to note that cross-validation was used only within the SVM hyperparameter tuning procedure, during the grid search on the training set. No cross-validation was performed for any other analyses or model training in this study. The rationale behind it is that Paddy Doctor dataset already provides a predefined and standardized train–test split released by the original authors. To ensure full comparability, we adhered to the official partitioning protocol. Consequently, all results reported here correspond to this fixed evaluation split, which guarantees methodological consistency with the benchmark configuration.

Compared with the original Paddy Doctor benchmarks, where CNN variants typically achieved accuracies in the range of 95–97% under full training sets, the results presented here demonstrate that contrastive dissimilarity can provide competitive results. The achieved accuracy of 98.2–98.4% underlines that the proposed method is robust in imbalanced and low-data scenarios (as originally proposed), and also capable of improving classification in well-resourced conditions. This observation extends the scope of contrastive dissimilarity, aligning with the insights of [6,7], who reported comparable robustness for music genre classification and for writer identification tasks across both balanced and imbalanced datasets.

While the proposed method does not yet achieve performance on par with the current state of the art, its primary contribution lies in introducing a novel perspective on feature representation and decision modeling. As widely recognized in the literature, methodological diversity is a key factor in extracting complementary information from the data, which is particularly valuable for the development of robust ensemble systems. Our approach demonstrates the ability to capture information that differs that those learned by standard pipelines, such as SVM classifiers trained on convolutional feature embeddings. These results suggest that the proposed framework represents a promising and functional direction, which can be further refined through future research, performance optimization, and systematic integration into ensemble strategies to enhance overall classification robustness and generalization.

Although the method shows encouraging performance, we identify two key practical limitations. First, the amount of data required for the approach to be advantageous is scenario-dependent, which makes it hard to define a single, clear cutoff where it consistently surpasses conventional baselines. Second, the framework relies on several tunable design choices that influence results and need calibration.

In addition, the Paddy Doctor dataset also highlights directions where complementary work would be valuable: for example, assessing robustness to field factors such as illumination changes, background clutter, and symptom severity, and validating transfer to unseen acquisition conditions such as different devices, seasons, or locations.

6. Conclusions

In this work, we investigated the use of contrastive dissimilarity for addressing class imbalance and limited data availability in rice disease diagnosis using the Paddy Doctor dataset. By integrating representation learning with a dissimilarity-based contrastive metric, the proposed approach successfully captured discriminative relationships between visually similar diseases, demonstrating competitive and, in some cases, superior performance compared to a baseline CNN model.

Our experiments revealed that contrastive dissimilarity not only maintained robustness under severe data scarcity (down to 5% of the training set) but also achieved the highest accuracy when trained with the full dataset (98.2–98.4%). This indicates that the method generalizes well across different levels of data availability, mitigating the adverse effects of class imbalance while preserving discriminative capability.

The results suggest that contrastive dissimilarity is a promising direction for precision agriculture, where data imbalance and limited annotations are common. Its ability to produce compact and effective embeddings while remaining resilient to uneven class distributions makes it suitable for broader applications in agricultural imaging and beyond, including other crops and disease diagnosis systems.

Future work will explore the integration of contrastive dissimilarity with transformer-based vision backbones and few-shot learning frameworks, aiming to enhance performance in ultra-low data regimes further. Additionally, the incorporation of multimodal information, such as spectral or environmental data, may provide complementary cues to improve diagnostic reliability in real-world field conditions further. Finally, we will evaluate cross-region, cross-season, and cross-dataset generalization under field acquisition conditions (lighting variation, occlusion, device shift) and will incorporate uncertainty calibration and open-set detection to reduce failures on unknown diseases.

Author Contributions: Conceptualization, F.P. and L.O.T.; methodology, F.P. and L.O.T.; validation, F.P. and L.O.T.; investigation, F.P. and L.O.T.; data curation, F.P. and L.O.T.; writing—original, F.P., L.O.T. and Y.M.G.C.; writing—review and editing, F.P., L.O.T., L.N. and Y.M.G.C.; supervision, L.O.T., L.N. and Y.M.G.C.; project administration, F.P., L.O.T., L.N. and Y.M.G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been partly supported by the National Council for Scientific and Technological Development (CNPq), Araucária Foundation—Grant 1144/2025, Governo do Estado do Paraná/SETI—Grant 1144/2025 and Coordination for the Improvement of Higher Education Personnel (CAPES)—Finance code 001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this article is available in IEEE DataPort at [https://dx.doi.org/10.21227/hz4v-af08], under Standard Access. Access requires an active IEEE DataPort subscription.

Acknowledgments: AI or AI-assisted tools were used in parts of the manuscript exclusively for language translation, language editing and grammar checking. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
BPB	Bacterial Panicle Blight

CD	Contrastive Dissimilarity
CDS	Contrastive Dissimilarity Space
CDV	Contrastive Dissimilarity Vector
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
RGB	Red, Green and Blue
SE	Squeeze and Excitation
SSDHR	Self-Supervised Deep Hierarchical Reconstruction
VGG	Visual Geometry Group
YOLO	You Only Look Once

References

1. Ahad, M.T.; Li, Y.; Song, B.; Bhuiyan, T. Comparison of CNN-based deep learning architectures for rice diseases classification. *Artif. Intell. Agric.* **2023**, *9*, 22–35. [CrossRef]
2. Zhang, J.; Huang, Y.; Pu, R.; Gonzalez-Moreno, P.; Yuan, L.; Wu, K.; Huang, W. Monitoring plant diseases and pests through remote sensing technology: A review. *Comput. Electron. Agric.* **2019**, *165*, 104943. [CrossRef]
3. Miftahushudur, T.; Sahin, H.M.; Grieve, B.; Yin, H. A survey of methods for addressing imbalance data problems in agriculture applications. *Remote Sens.* **2025**, *17*, 454. [CrossRef]
4. Bruno, A.; Bhatt, C.; Aoun, N.B.; Malaviya, P.; Mulla, A. Deep Learning Techniques for Accurate Classification of Rice Diseases: A Comprehensive Study. In *Proceedings of the Intelligent Systems Conference*; Springer: Berlin/Heidelberg, Germany , 2024; pp. 452–470.
5. Teixeira, L.O.; Bertolini, D.; Oliveira, L.S.; Cavalcanti, G.D.; Costa, Y.M. Contrastive dissimilarity: Optimizing performance on imbalanced and limited data sets. *Neural Comput. Appl.* **2024**, *36*, 20439–20456. [CrossRef]
6. Costanzi, G.H.; Teixeira, L.O.; Felipe, G.Z.; Cavalcanti, G.D.; Costa, Y.M. Music Genre Classification Using Contrastive Dissimilarity. In *Proceedings of the 2024 31st International Conference on Systems, Signals and Image Processing (IWSSIP)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–8.
7. Pignelli, F.; Teixeira, L.O.; Bertolini, D.; Nanni, L.; Cavalcanti, G.D.; Costa, Y.M. Contrastive Dissimilarity for Writer Identification: A proof of concept. In *Proceedings of the 2025 32nd International Conference on Systems, Signals and Image Processing (IWSSIP)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–5.
8. Teixeira, L.O.; Bertolini, D.; Oliveira, L.S.; Cavalcanti, G.D.; Costa, Y.M. Improving open set recognition with dissimilarity-based metric learning. *Knowl.-Based Syst.* **2025**, *327*, 114108. [CrossRef]
9. Petchiammal, A.; Briskline Kiruba, B.; Murugan, D.; Arjunan, P. Paddy doctor: A visual image dataset for automated paddy disease classification and benchmarking. In *Proceedings of the 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD)*; Association for Computing Machinery: New York, NY, USA , 2023; pp. 203–207.
10. Zhang, Z.; Wang, R.; Huang, S. Multi-Model Identification of Rice Leaf Diseases Based on CEL-DL-Bagging. *AgriEngineering* **2025**, *7*, 255 . [CrossRef]
11. Mookkandi, K.; Nath, M.K.; Dash, S.S.; Mishra, M.; Blange, R. A Robust Lightweight Vision Transformer for Classification of Crop Diseases. *AgriEngineering* **2025**, *7*, 268. [CrossRef]
12. Petchiammal, A.; Murugan, D. Automated Paddy Leaf Disease Identification using Visual Leaf Images based on Nine Pre-trained Models Approach. *Procedia Comput. Sci.* **2025**, *252*, 118–126. [CrossRef]
13. Padhi, J.; Korada, L.; Dash, A.; Sethy, P.K.; Behera, S.K.; Nanthaamornphong, A. Paddy leaf disease classification using EfficientNet B4 with compound scaling and swish activation: A deep learning approach. *IEEE Access* **2024**, *12*, 126426–126437. [CrossRef]
14. Thanuboddi, N.; Nelakuditi, U.R. Hybrid deep learning for smart paddy disease diagnosis using self supervised hierarchical reconstruction and attention based temporal analysis. *Sci. Rep.* **2025**, *15*, 34917. [CrossRef] [PubMed]
15. Bera, A.; Bhattacharjee, D.; Krejcar, O. An attention-based deep network for plant disease classification. *Mach. Graph. Vis.* **2024**, *33*, 47–67. [CrossRef]
16. Pekalska, E.; Duin, R.P. Dissimilarity-based classification for vectorial representations. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*; IEEE: Piscataway, NJ, USA, 2006; Volume 3; pp. 137–140.
17. Costa, Y.M.; Bertolini, D.; Britto, A.S., Jr.; Cavalcanti, G.D.; Oliveira, L.E. The dissimilarity approach: A review. *Artif. Intell. Rev.* **2020**, *53*, 2783–2808. [CrossRef]
18. Kulis, B. Metric learning: A survey. *Found. Trends Mach. Learn.* **2013**, *5*, 287–364. [CrossRef]
19. Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; Gong, Y. Locality-constrained Linear Coding for image classification. In *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2010; pp. 3360–3367. [CrossRef]

20. Tan, M.; Le, Q.V. EfficientNetV2: Smaller Models and Faster Training. In Proceedings of the 38th International Conference on Machine Learning, Virtual, 18–24 July, 2021; Volume 139, pp. 10096–10106. Available online: https://proceedings.mlr.press/v139/tan21a.html (accessed on 1 November 2025).
21. Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst. Appl.* **2017**, *73*, 220–239. [CrossRef]
22. Kuncheva, L.I.; Whitaker, C.J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach. Learn.* **2003**, *51*, 181–207. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.