



Article

# Soybean Canopy Stress Classification Using 3D Point Cloud Data

Therin J. Young <sup>1</sup>, Shivani Chiranjeevi <sup>1</sup>, Dinakaran Elango <sup>2</sup> , Soumik Sarkar <sup>1,3</sup>, Asheesh K. Singh <sup>2</sup>, Arti Singh <sup>2,\*</sup>, Baskar Ganapathysubramanian <sup>1,3,\*</sup> and Talukder Z. Jubery <sup>3,\*</sup>

<sup>1</sup> Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA; theriny@iastate.edu (T.J.Y.); shichir@iastate.edu (S.C.); soumiks@iastate.edu (S.S.)

<sup>2</sup> Department of Agronomy, Iowa State University, Ames, IA 50011, USA; delango@iastate.edu (D.E.); singhak@iastate.edu (A.K.S.)

<sup>3</sup> Translational AI Research and Education Center, Iowa State University, Ames, IA 50011, USA

\* Correspondence: arti@iastate.edu (A.S.); baskarg@iastate.edu (B.G.); znjubery@iastate.edu (T.Z.J.)

**Abstract:** Automated canopy stress classification for field crops has traditionally relied on single-perspective, two-dimensional (2D) photographs, usually obtained through top-view imaging using unmanned aerial vehicles (UAVs). However, this approach may fail to capture the full extent of plant stress symptoms, which can manifest throughout the canopy. Recent advancements in LiDAR technologies have enabled the acquisition of high-resolution 3D point cloud data for the entire canopy, offering new possibilities for more accurate plant stress identification and rating. This study explores the potential of leveraging 3D point cloud data for improved plant stress assessment. We utilized a dataset of RGB 3D point clouds of 700 soybean plants from a diversity panel exposed to iron deficiency chlorosis (IDC) stress. From this unique set of 700 canopies exhibiting varying levels of IDC, we extracted several representations, including (a) handcrafted IDC symptom-specific features, (b) canopy fingerprints, and (c) latent feature-based features. Subsequently, we trained several classification models to predict plant stress severity using these representations. We exhaustively investigated several stress representations and model combinations for the 3-D data. We also compared the performance of these classification models against similar models that are only trained using the associated top-view 2D RGB image for each plant. Among the feature-model combinations tested, the 3D canopy fingerprint features trained with a support vector machine yielded the best performance, achieving higher classification accuracy than the best-performing model based on 2D data built using convolutional neural networks. Our findings demonstrate the utility of color canopy fingerprinting and underscore the importance of considering 3D data to assess plant stress in agricultural applications.

**Keywords:** iron deficiency chlorosis; terrestrial laser scanning; virtual phenotyping; canopy fingerprints



**Citation:** Young, T.J.; Chiranjeevi, S.; Elango, D.; Sarkar, S.; Singh, A.K.; Singh, A.; Ganapathysubramanian, B.; Jubery, T.Z. Soybean Canopy Stress Classification Using 3D Point Cloud Data. *Agronomy* **2024**, *16*, 1181. <https://doi.org/10.3390/agronomy14061181>

Academic Editor: Yanbo Huang

Received: 8 April 2024

Revised: 16 May 2024

Accepted: 25 May 2024

Published: 30 May 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Soybean (*Glycine max* (L.) Merr.) is one of the most important crops globally, with significant nutritional and economic impact. It is a powerhouse of essential nutrients that offers a complete protein profile consisting of vital vitamins and minerals, making it indispensable in vegetarian and vegan diets. Beyond direct human consumption, soybeans play a pivotal role in the industry, from serving as a primary ingredient in many processed foods to their transformation into biodiesel [1]. Moreover, a large fraction of the soybean harvest worldwide is utilized as high-protein livestock feed, underpinning the poultry, swine, and aquaculture sectors.

Plant stress is a state of plant growth under non-ideal environmental conditions caused by various biotic (pathogen, insect, pest, and weed) and abiotic (temperature stress, nutrient deficiency, toxicity, herbicide) factors [2]. It is a constant threat to global food security, and the ongoing challenge is to prevent and protect against yield losses due to

these stresses. Iron Deficiency Chlorosis (IDC) in soybeans is an abiotic stress that can significantly impact plant health and yield. When soybean plants suffer from IDC, they display characteristic interveinal chlorosis in young leaves, with leaf tissue yellowing while the veins stay green [3]. The disorder predominantly arises in high pH, calcareous soils that impede iron's bioavailability to the plant, even if the soil's overall iron content is abundant. Froehlich et al. [4] reported a robust linear association between soybean yield reductions and both visual chlorosis ratings and stunted growth. Specifically, each increment in chlorosis rating corresponded to a 90% increase in average yield loss. Additionally, for every 1% decrease in plant height, a yield reduction of 1.6% was observed. Interestingly, increased chlorosis severity led to decreased lodging due to shorter plant stature and also delayed maturity [4]. Peiffer et al. [5] estimated that the annual yield loss to farmers due to IDC in the USA alone is around \$260 million. The substantial yield reductions attributed to IDC can result in significant economic losses for farmers, underlining the need for effective management strategies, including cultivating IDC-tolerant soybean varieties. Addressing IDC is critical not only for the agronomic and economic ramifications but also to ensure the nutritional quality of the soybean crop. The conventional method of evaluating IDC stress symptoms is performed manually by visually evaluating the stress symptoms across the canopy. However, this approach is subjective, time-consuming, and prone to inter- and intra-rater inaccuracies [2]. Therefore, the past decade has seen multiple efforts toward developing a high-throughput phenotyping system for automated IDC rating using machine learning.

To phenotype soybean canopy biotic or abiotic stresses, various 2D imaging modalities, including RGB [6,7], hyperspectral [8–10], and thermal images [9,10] are used. Among these modalities, RGB imaging has been extensively employed for IDC symptom detection [11–14]. For instance, Bai et al. [11] investigated the utility of RGB images captured under field conditions for IDC scoring, achieving an accuracy of 81%. In the same way, Dobbels et al. [14] used UAS with an RGB camera to improve screening in the field for soybean IDC tolerance, obtaining an accuracy of 77%. In another study, Hassnijalilian et al. [13] utilized ground-based RGB images of plots and achieved an F1 score accuracy of 0.78. Image-based phenotyping systems are also packaged as smartphone apps designed to classify plant stresses in field environments [12]. However, while 2D methods offer reliability and versatility, they struggle to capture stress expression in the lower parts of the plant canopy.

To capture the full canopy, including the lower parts, LiDAR-equipped 3D scanners have been added to modern phenotyping systems [15–18]. These scanners capture comprehensive, three-dimensional point clouds of the canopy, including the complete lower part of the canopy, providing improved input features for stress classification algorithms and enhancing classification models. Though more expensive than image-based 3D reconstructions such as structure-from-motion (SFM) [19–21], LiDAR systems have provided more accurate measurements of plant phenotypes such as height and biomass as a result of their finer spatial resolution compared to SFM [22]. LiDAR systems were utilized to estimate structural traits including height [21,23,24], leaf area index [25–27], biomass [19,20] and biotic or abiotic stresses that are expressed mostly via a change in the structural traits [15,28–30]. LiDAR-based high-throughput phenotyping systems have been deployed for situations where the stress is expressed via the change in canopy color and have primarily been limited to applications in controlled environments. Combining three-dimensional point clouds of the canopy with RGB and hyperspectral information has shown promise in limited indoor settings. Studies have explored the use of multimodal 3D data for assessing abiotic and biotic stress, including IDC, in controlled environments [15,31–34]. Anesley et al. [34] employed LiDAR in a greenhouse setting to identify drought tolerance markers in potato breeding. They leveraged LiDAR data to estimate plant height, 3D leaf area, leaf area projection on the ground, and leaf angle to assess drought response across 64 genotypes. Similarly, Khanna et al. [31] developed a pipeline that employed reconstructed 3D hyperspectral images alongside 2D color and infrared stereo images to predict water, nitrogen, and weed stress on greenhouse sugarbeets. Ríos-Toledo et al. [35] utilized RGB and 3D

point cloud data to analyze various abiotic stressors, highlighting the superiority of 3D over 2D imaging in stress prediction. However, these studies are proof-of-concept, and despite their potential, most 3D stress-related multi-modal phenotyping systems remain confined to controlled environments, posing limitations for field-based experiments.

Extracting meaningful representations from large field-based datasets is a bottleneck in phenotyping. Studies have underscored the importance of feature representation [36,37]. While traditional models typically rely on handcrafted representation features, promising results have been demonstrated with canopy fingerprints [38–41] and latent features [3,42]. Previous studies have utilized these features for soybean stresses but focused exclusively on 2D images. In this research, we aim to leverage the potential of 3D point clouds with RGB color to enhance stress classification algorithms and significantly improve the understanding and detection of IDC stress in soybean plants.

The main goal of this study is to explore different ways of measuring IDC stress severity in soybean plants, such as fingerprints, hand-crafted features, and latent features, using both 3D point cloud data and 2D image data. We aim to train multiple classification models using these representations and conduct a comprehensive performance comparison to generate insights into the usage of 3D data for stress phenotyping. Additionally, we want to conduct experiments to assess the impact of data balancing on model training. Our contributions to this work are as follows:

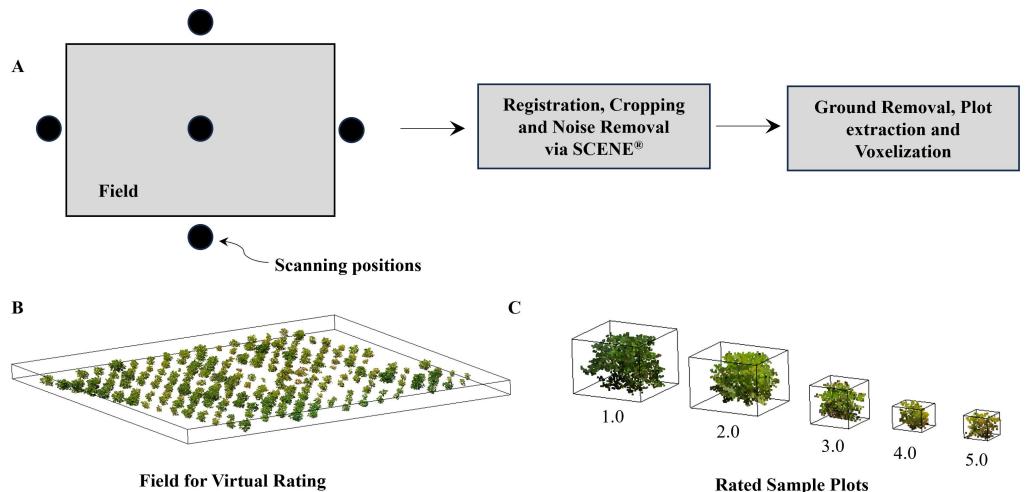
- Developed an in-field scanning and automated plot extraction process for high-throughput phenotyping of soybean canopies.
- Generated handcrafted color and fingerprint features from 3D point cloud for each extracted plot.
- Explored the impact of different feature representations on the classification results, showing that 3D handcrafted features exhibited superior class separation compared to 2D handcrafted features. Achieved 95% and 97% classification accuracy for 3D handcrafted and fingerprint features, respectively.

## 2. Materials and Methods

### 2.1. Location and Field Scanning

The data for this study were collected from field experiments performed in iron-deficient soil at Iowa State University's Agricultural Engineering/Agronomy Research Farm, IA, USA (latitude of 42.025, longitude of −93.778). The soil type is Harps Series with a pH range of 7.5 to 8. The field experiment consisted of soybean cultivars from multiple countries with diverse maturities, seed weights, and stem terminations. We planted 109 cultivars with two checks (Clark and Iso-Clark). All the entries were replicated twice. In May 2021, the cultivars were hand-planted in hill plots, with three seeds per hill and 0.76 m spacing between each hill. Each plot consisted of a single, unbounded hill. No plants were thinned, and weeding was conducted manually as needed. The entire field measures 40 by 40 feet, which is equivalent to 0.04 acres.

Beginning on 6 July 2021, point cloud scans of the field were captured twice a week, up until 2 August 2021, using a Faro Focus S350 Terrestrial Laser Scanner (TLS). The scanner can produce dense point clouds (2 mm (about 0.08 in) resolution) and capture RGB images with up to 70 MP resolution via a built-in camera. The scanner was mounted on a 4-foot tripod, and point cloud scanning was performed from four sides and the center of the field to capture the entire canopy, as shown in Figure 1A. The field scanning period occurred during the V4–V5 and R1 growth stages of the plants.



**Figure 1.** (A) Scanning and Plot Extraction Process: The illustration shows the field image with scanning positions, covering five distinct locations, including four sides and near the center of the field. After capturing the point cloud data, we cleaned, processed, and registered it using Faro SCENE 2021® software. Subsequently, our in-house Python script was utilized to extract individual plots from the registered point cloud data and remove the ground [43] (B) Rating: Following the removal of the ground portion, all plots were visualized using CloudCompare 2.11.3®, and an expert assessed each plot's IDC stress on a scale of 1 to 5. (C) Visualization: Representative canopies for each IDC stress rating.

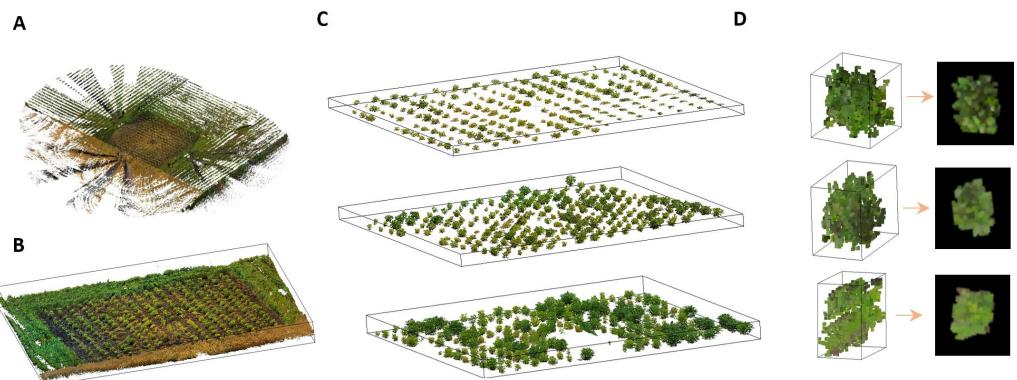
## 2.2. Plot Extraction and IDC Rating

Using the Faro SCENE 2021® software [44], we performed color mapping and registration of the five-point clouds captured from the field. Following the registration process, we conducted post-processing on the registered point cloud. This involved removing noise and achieving point uniformity through voxelization. The color of the voxel is the average of all the RGB values for the points within a voxel, as described in Open3D's user documentation [45]. We segmented ground points from canopy points and saved each canopy as a separate file. We cannot directly segment plant canopies by color because plants and soil often appear identical near the ground. To overcome this, we first create a rough separation. We fit a plane to the entire point cloud data, keeping everything above it as a preliminary canopy. Next, we segment individual plots using the connected component algorithm and estimate their centers. Finally, we create a refined ground plane for each plot based on a square area around each plant center. With this more accurate ground definition, we fit a final plane through these refined ground points and keep everything above it as the true plant canopy. We utilized an in-house Python script for post-processing and plot extraction, as described in our previous study by Young et al. [43]. Out of the 1100 canopies extracted, we selected 700 for downstream analysis. The selection process excluded canopies with poor color and structural quality, primarily caused by overcast conditions (80%), incompleteness (15%), or noise (5%). Incompleteness refers to parts of the canopy not visible to the scanner during scanning, while noise results from canopy movement due to wind.

Expert raters virtually conducted IDC ratings of the canopies by carefully examining each point cloud [46]. These virtual visual ratings (VVR) used a scale of 1–5 following the traditional visual ratings scale [12], with 5 indicating the highest disease severity. IDC, stemming from higher soil pH reducing iron solubility, reduced iron intake by leaves and resulted in chlorosis. Stress rating 1 signified the absence of chlorosis, with canopy leaves retaining their green color. Rating 2 meant that the upper leaves were slightly yellowing; rating 3 meant that the upper leaves had interveinal chlorosis; rating 4 meant that the leaves had interveinal chlorosis and growth was slowed down and finally, rating 5 meant

that the leaves had extreme chlorosis, growth was slowed down, and necrosis appeared on new leaves.

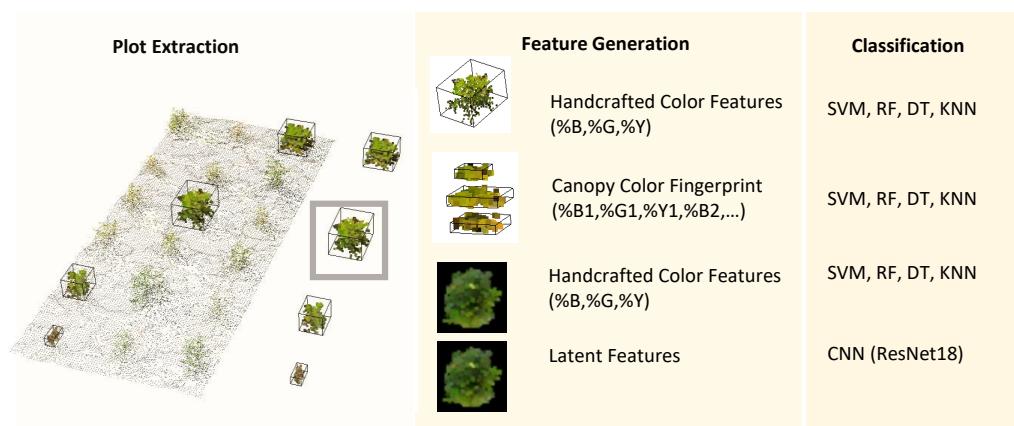
We also generated 2D top-view projections of cleaned voxelized 3D plots of each hill plot. Each projection was saved as a PNG file. The script for generating these 2D images from the 3D voxelized point cloud is accessible on our GitHub repository: [https://github.com/znjubery/3D\\_Plant\\_Stress\\_Repo\\_ty](https://github.com/znjubery/3D_Plant_Stress_Repo_ty) (accessed on 24 May 2024). A schematic of the complete data collection and curation workflow is shown in Figure 1. Figure 2 presents a selection of images showcasing the 3D point clouds at different stages of the data curation process. It includes 3D point clouds of the entire field, individual plots, and the corresponding top-view projected 2D images.



**Figure 2.** A selection of images showcasing the 3D point clouds at various stages of the data curation process: (A) Original registered RGB point cloud. (B) Cropped point cloud focused on the field. (C) Point clouds of all plots, with ground removed, captured at three different time points; note that some plots were excluded due to overcast conditions, incompleteness, or noise. (D) Sample 3D point cloud and the corresponding top-view 2D projection.

### 2.3. Feature Generation

We utilize three distinct methods for generating features from 3D point clouds and 2D images, as shown in Figure 3. These approaches encompass handcrafted features, canopy fingerprint features, and latent features. The first two methods involve explicit feature extraction, while latent features are generated implicitly during the construction of deep-learning classification models. Latent features are implicitly generated during neural network training and are not manually extracted or visualized. In contrast, the handcrafted features are manually extracted based on predefined criteria.



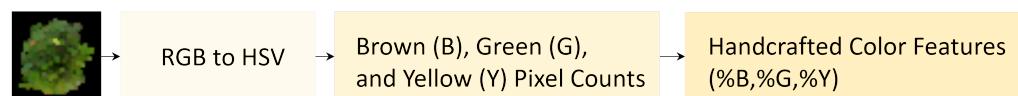
**Figure 3.** Feature Generation and Classification: After plot extraction, we performed feature generation and classification. To showcase the diversity in canopy color and size present in the dataset,

we increased the point density of four extracted plots. For each plot, we extracted handcrafted color and canopy fingerprint features from the 3D point cloud. From the 2D top-view projection image, we extracted handcrafted color features and generated latent features using a neural network. These features, combined with the IDC ratings, were used to train various classification algorithms, including Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest (RF), and Convolutional Neural Network (CNN), to predict the IDC class labels.

### 2.3.1. Handcrafted Features for 3D Point Clouds and 2D Images

The creation of handcrafted features is motivated by the expert's reasoning applied during virtual assessments. The expert examined color signatures from green to yellow to brown, specifically assessing the degree of discoloration originating from the green shade. This analysis encompassed the transition from chlorosis (yellowing) to necrosis (browning).

To mimic the rater's approach of quantifying color signatures, we used Python with OpenCV and Open3D to process the RGB 3D point cloud or 2D image of the canopy into HSV (Hue, Saturation, Value) values, as shown in Figure 4. For 2D images, we read the image using OpenCV's 'cv2.imread()' function. For 3D point clouds, we utilized Open3D's functionalities to load and manipulate the data.



**Figure 4.** Overview of Handcrafted feature generation from 2D images. We used a similar workflow for Handcrafted feature generation from 3D point clouds.

First, we extracted the RGB values from the 3D point cloud and converted them to HSV using OpenCV's 'cv2.cvtColor()' function. For 2D images, we directly converted the image from RGB to HSV. We defined the hue ranges for green, yellow, and brown colors as 45–64, 31–45, and 0–31, respectively. Using these predefined hue ranges, we created masks for each color with OpenCV's 'cv2.inRange()' function to identify the pixels or points within the specified hue range.

Finally, we calculated the percentage of green (%G), yellow (%Y), and brown (%B) colors in each plot by dividing the number of points or pixels of each color by the total number of points or pixels. These %G, %Y, and %B percentages are considered handcrafted features. Such handcrafted features have proven successful in automated IDC ratings of 2D images [12], and this work extends the approach to 3D point clouds.

### 2.3.2. Canopy Fingerprinting for 3D Point Clouds

Canopy fingerprinting [43] is an approach for characterizing the three-dimensional structure of plant canopies. This approach involves analyzing 3D point cloud data of canopies to create detailed yet interpretable phenotypic traits. These fingerprints are generated by dividing the canopy into smaller, sub-canopy-scale components and extracting specific geometric features. This method is particularly useful for identifying patterns, querying similar canopies, or recognizing canopies with particular shapes. The features extracted using this method can feed into statistical or machine learning techniques to map to various traits.

In the current study, we extend the canopy fingerprinting concept to extract not only sub-canopy scale geometric features but also the sub-canopy scale color features that are representative of IDC stress. This process involves segmenting a canopy into a series of  $2n + 1$  equally divided sections, referred to as sub-canopies, to capture the disease symptoms within each sub-canopy. The division is conducted vertically along the height direction, yielding the  $2n + 1$  sections. Each sub-canopy is then assigned a color signature indicating the percentage of each of three colors: green, yellow, and brown ([%G, %Y, %B]). Essentially, instead of capturing the canopy scale amount of chlorosis and necrosis

(as performed in Section 2.3.1), here, the vertical distribution of chlorosis and necrosis is captured, thus providing a finer scale of detail on stress progression. The final fingerprint for a canopy's 3D RGB point cloud is represented by concatenating the color signatures of the sub-canopies into a single feature vector of size  $[(2n + 1) \times 3, 1]$ .

#### 2.4. Classification Models

We used several machine learning models to map the generated features to the visually rated IDC ratings and construct classification models. The virtual rating served as the categorical output variable (classes), while the inputs were the feature vectors. These classification models are then utilized to predict IDC ratings for different feature vectors. We train decision trees (DT), random forests (RF), k-nearest neighbors (KNN), and support vector machines (SVM) using both handcrafted and fingerprint feature vectors. The handcrafted feature vector size was 3, while the fingerprint feature vector size was 9.

Decision trees construct predictive models using a tree-like structure to categorize observations [47]. For the DT classifier, the splitting was performed based on the Gini impurity; the max depth and random state were set to 4 and 0, respectively. The maximum depth of the decision tree is an integer that controls the overall complexity of the tree and must not exceed the number of samples used to train the model. Too large a maximum depth can cause over-fitting of the model [48]. The random state parameter controls the randomness of the estimator, resulting in randomly permuted features at each decision tree split. RF is an ensemble method that addresses over-fitting by combining multiple decision trees [49]. We utilized 100 decision trees with Gini impurity as splitting conditions for the RF implementation.

KNN assigns class labels based on similarity metrics in the feature space [50]. For the KNN classifier, four neighbors were used for the similarity estimation. Neighbors are selected based on Euclidean distance, and all points are equally weighted when a new point is queried. SVM is a widely used supervised technique that employs kernel functions to separate data into higher-dimensional spaces [51]. We utilized the Radial Basis Function (RBF) kernel, known for its effectiveness with non-linear and high-dimensional data. To balance smoothness and accuracy in the decision boundary, we employ a regularization parameter set to 1. This parameter interacts with another key factor, gamma, which controls the influence of individual training points. A small gamma value creates a smoother boundary by extending each point's influence, while a large gamma focuses on nearby points, potentially leading to overfitting. The algorithm automatically sets gamma based on feature statistics to ensure consistent impact across datasets with varying feature scales.

We conducted training on ResNet18 for 2D image classification tasks using a fine-tuning approach on a ResNet18 pre-trained on ImageNet. The process involved training only with the final four layers after adding a new classification head. To standardize the images, we resized them to 224 by 224 pixels and normalized the color values based on mean and standard deviation. We used random rotations to augment the data. The model was trained using the cross-entropy loss function and the Adam optimizer for 100 epochs, with a batch size of 30, and a learning rate set to 0.001.

Because of this study's limited number of samples, achieving an equivalent distribution within the training and test data sets after random splitting proved challenging. It was necessary to carefully split the data manually or semi-automatically to ensure that both datasets had representative samples regarding class labels and intra-class variability. Hence, we opted for K-fold cross-validation, elaborated in the Section 2.4.1.

##### 2.4.1. Evaluation Metrics

The classifier's effectiveness can be evaluated using a confusion matrix. In this matrix, the diagonal elements represent the number of observations where the predicted rating matches the actual rating, while the off-diagonal elements indicate misclassified observations.

An example confusion matrix for a binary classification problem is shown below:

|                      | Predicted Positive (TP) | Predicted Negative (TN) |
|----------------------|-------------------------|-------------------------|
| Actual Positive (TP) | TP                      | FN                      |
| Actual Negative (TN) | FP                      | TN                      |

In our analysis, we evaluated the performance of the models using information extracted from the confusion matrix. We used accuracy, mean per-class accuracy, and misclassification as key metrics to compare the effectiveness of our models.

Accuracy provides insight into how well the model predicts the dataset's correct outcomes. It is calculated using the formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

Mean per-class accuracy offers a refined understanding of how the classifier performs for individual classes. Per-class accuracy considers the fraction of correctly predicted instances for each class, which is especially valuable when dealing with imbalanced classes. The formula for per-class accuracy for each class ' $i$ ' is as follows:

$$\text{Per class accuracy} = \frac{\text{i-th observation of row } i}{\text{Sum of observations of row } i}, \quad i = 1, \dots, n$$

where ' $n$ ' is the number of classes, and 'row  $i$ ' represents the  $i$ -th row in the confusion matrix.

Mean per-class accuracy (MPCA) is the average of per-class accuracy across all classes and can be calculated as:

$$\text{Mean Per Class Accuracy} = \frac{1}{n} \sum_{i=1}^n \text{Per class accuracy}$$

Additionally, we utilized misclassification costs to quantify the impact of misclassification errors. This was achieved through a misclassification cost matrix ( $w_{ij}$ ) detailing the costs associated with wrongly predicted pairs of actual and predicted ratings, as outlined in Table 1. In this matrix, the off-diagonal elements signify the costs of misclassification for each rating, typically represented by finite real values. For instance, if an observation's actual rating is 5, the cost of misclassifying it as rating 1 is four times higher than misclassifying it as rating 4, and so forth. The misclassification cost can be computed using the formula:

$$\text{Misclassification Cost} = \sum_{i=1}^n \sum_{j=1}^n \frac{CM_{ij} \times w_{ij}}{N}$$

**Table 1.** Cost matrix  $w_{ij}$ .

| Actual | Predicted |   |   |   |   |
|--------|-----------|---|---|---|---|
|        | 0         | 1 | 2 | 3 | 4 |
| 0      | 0         | 1 | 2 | 3 | 4 |
| 1      | 1         | 0 | 1 | 2 | 3 |
| 2      | 2         | 1 | 0 | 1 | 2 |
| 3      | 3         | 2 | 1 | 0 | 1 |
| 4      | 4         | 3 | 2 | 1 | 0 |

We then use cross-validation to estimate how well each classifier performs on average. Cross-validation is a technique to test how accurate a model is with new data. Cross-validation is important because considering only accuracy can be misleading due to biases and over-fitting. By considering bias and variance, cross-validation helps us understand how well a model can generalize to new data. We use a common approach called k-fold cross-validation with a value of k equal to 5, which strikes a good balance between bias and variance [52]. We repeat this process five times to calculate the average misclassification

error for each model. While accuracy and mean per-class accuracy show how well the model does on the same dataset, the average cross-validation misclassification error tells us how well the model performs on different datasets.

#### 2.4.2. Data Imbalance

In our study, the dataset exhibited an imbalance in distribution across IDC classes, primarily with most observations falling into IDC class 5 (VVR = 5) and fewer observations belonging to class 1 (VVR = 1). This imbalance arises from the sample of soybean accessions and soil conditions that promoted the expression of IDC symptoms. This resulted in a higher count for class 5 than for class 1. Additionally, the other classes (VVR = 2/3/4) displayed a relatively balanced distribution among themselves. The class imbalance within the dataset is visualized in Table 2.

**Table 2.** Distribution of Samples Based on IDC Rating in Our Dataset.

| IDC Rating | 1  | 2   | 3   | 4   | 5   |
|------------|----|-----|-----|-----|-----|
| Counts     | 86 | 124 | 123 | 164 | 226 |

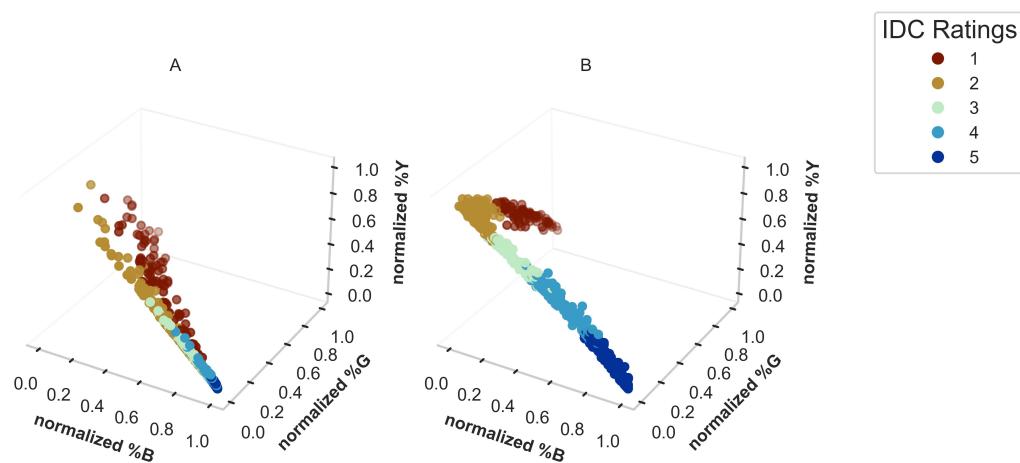
To address the class imbalance, we utilized the synthetic minority over-sampling technique (SMOTE) [53,54], a data augmentation method. SMOTE generates synthetic samples resembling existing minority-class data points. It selects a minority data point and its  $k$  nearest neighbors then creates new samples through random interpolation. This helps counter class imbalance effects on the classifier. We incorporated an advanced variant, SVM-SMOTE, which uses support vector machine principles to enhance synthetic sample generation. Instead of random selection, SVM-SMOTE identifies key support vectors in the minority class to strategically create samples closer to decision boundaries. This approach balances classes and refines class boundaries, potentially boosting the classifier's overall performance and generalization on imbalanced datasets.

### 3. Results and Discussion

#### 3.1. Handcrafted Stress Representation Comparisons

The scatter plots presented in Figure 5 offer a comprehensive understanding of the intricate relationship between the three color percentages extracted from the complete canopy point cloud data. Each data point featured on the plot corresponds to a unique soybean canopy and is color-coded to reflect the canopy's IDC class rating. Significantly, a discernible demarcation between a canopy's IDC severity class rating and its color composition is observable. This observation emphasizes the efficacy of employing color percentages as an informative feature for classifying IDC severity.

Furthermore, the distinction between a canopy's IDC severity class rating and its color composition is particularly pronounced when analyzing the handcrafted features derived from 3D point clouds (Figure 5A) instead of those extracted from 2D images (Figure 5B). This enhanced separation underscores the distinct advantages of leveraging the additional dimensionality and complexity offered by 3D point cloud data. A similar overlap in IDC severity class ratings within the handcrafted features extracted from 2D images was reported by Naik et al. [12]. The discrepancy, or more overlapping in 2D handcrafted features than 3D features, is understandable. Visual rating and 3D handcrafted features involve observing the entire canopy, whereas 2D handcrafted features were extracted from partial canopy views, thus utilizing less comprehensive data. Additionally, for severity ratings reliant on color percentage, the invisibility of any part of the diseased plant's canopy (containing healthy or diseased leaves) will alter the percentage composition.

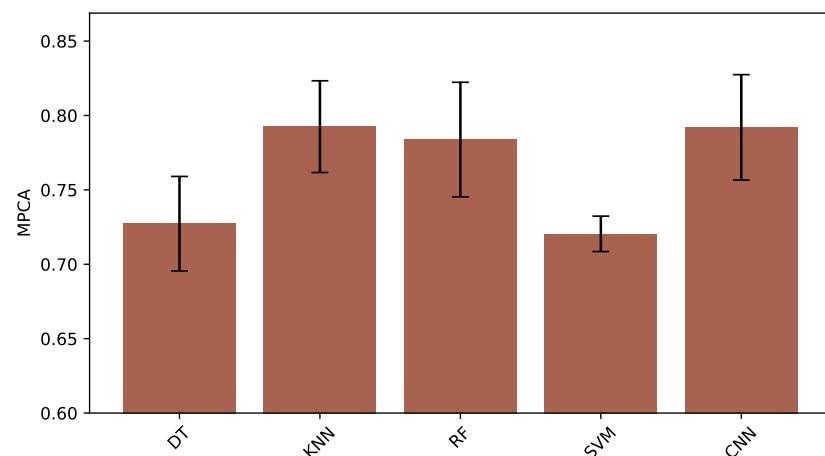


**Figure 5.** Comparing (A) 2D and (B) 3D Handcrafted Features: Rating-Specific Distribution. Each point on the plot represents the handcrafted color features of an IDC-stressed soybean canopy. Notice that the 3D handcrafted features exhibit superior class separation compared to their 2D counterparts.

### 3.2. Comparing How Various Representations Affect Classification Results

We generated handcrafted, fingerprint, and latent representations. We trained DT, KNN, RF, SVM, and ResNET18 models and compared their performance using mean per class accuracy (MPCA).

From 2D images, we generated a handcrafted stress representation and trained different classifiers like DT, KNN, RF and SVM to map the handcrafted features to their respective stress rating. We also used a ResNet18 CNN to directly map the learned latent features of RGB images to stress ratings. Figure 6 shows that after training on a balanced dataset Figure S1 using SVMSMOTE, the models' average per-class accuracy ranged from 0.7 to 0.84, with the ResNet18 CNN having the best performance.

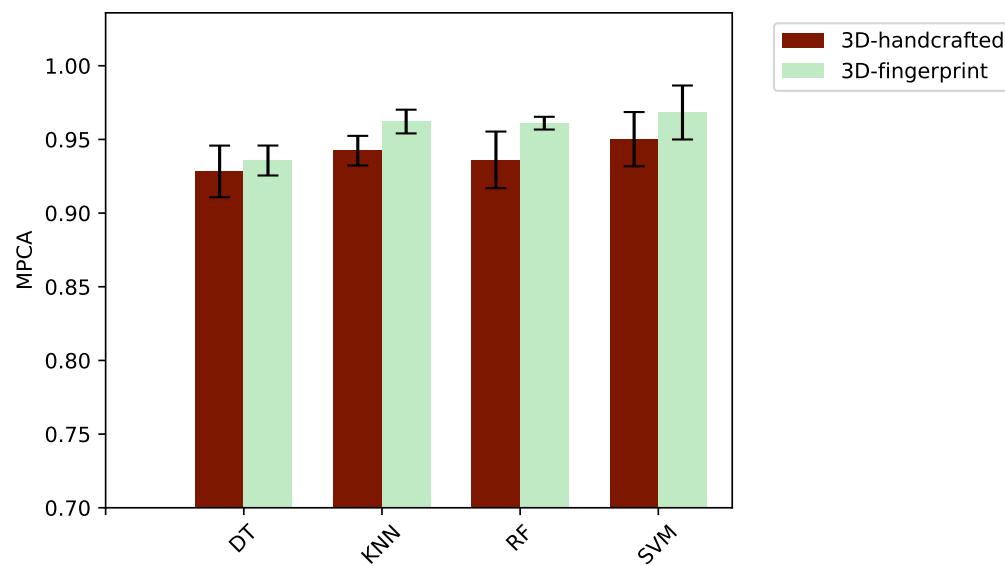


**Figure 6.** Comparing Classification Accuracy for Different Models on 2D Data: The evaluated models encompass 2D Handcrafted Color Features and 2D CNN latent features. The x-axis represents the various models (DT, KNN, RF, SVM, and CNN), and the bar plot illustrates the 5-fold accuracies achieved by each model.

From the 3D RGB point cloud, we generated handcrafted features and fingerprint stress representations to train different classifiers like DT, KNN, RF, and SVM. Figure 7 illustrates the classifiers' performance on the balanced training dataset using SVMSMOTE. The classifier based on handcrafted features had an average per-class accuracy varying from 0.93 to 0.95, while models trained with fingerprint features resulted in accuracy ranging

from 0.94 to 0.97. The SVM classifier trained on the 3D fingerprint representation returned the highest mean per-class accuracy.

The disparity in performance observed among classifiers trained on handcrafted features and fingerprint stress representations can be attributed to several key factors: (a) Fingerprint stress representations offer a more intricate and discriminative feature set than handcrafted features extracted directly from the RGB point cloud. This richer feature representation enables classifiers to capture finer patterns and nuances in the data, improving classification accuracy. (b) The higher dimensionality of fingerprint features allows for a more comprehensive representation of the underlying data distribution, facilitating the learning of complex decision boundaries. (c) The suitability of different classifiers for handling specific characteristics of the feature representations plays a crucial role. For instance, Support Vector Machines (SVMs) are well-known for their effectiveness in handling high-dimensional data and nonlinear decision boundaries, making them particularly adept at exploiting the discriminative power of fingerprint features. (d) Techniques like SVMSMOTE for handling data imbalance can also improve classifier performance by providing a more balanced representation of the data during training. Overall, the combination of richer feature representations, higher dimensionality, appropriate classifier selection, and effective data balancing techniques collectively contribute to the superior performance observed in the SVM classifier trained on the 3D fingerprint representation compared to other classifiers trained on handcrafted features.



**Figure 7.** Comparing Classification Accuracy for Different Models on 3D Data: The evaluated models encompass 3D Handcrafted Color Features and 3D Fingerprints. The x-axis represents the various models (DT, KNN, RF, and SVM), and the bar plot illustrates the five-fold accuracy achieved by each model.

### 3.3. Addressing Data Imbalance

In Table 3, we present the notable impact of our data balancing efforts on the outcomes of our classification models, specifically focusing on the performance of models trained on imbalanced and balanced 2D handcrafted features. The table highlights the mean per-class accuracy (MPCA) attained through 5-fold cross-validation and the corresponding standard deviations in parentheses. For models like k-nearest neighbors (KNN) with four neighbors and support vector machine (SVM) employing the radial basis function with a regularization parameter of 1, as well as decision trees constrained by a maximum depth of 4, there is a discernible increase in MPCA after implementing data balancing techniques. We note that all models have small standard deviations, suggesting robust models. Additionally, we observe a reduction in the decision tree and SVM classifier's standard deviation, indicating

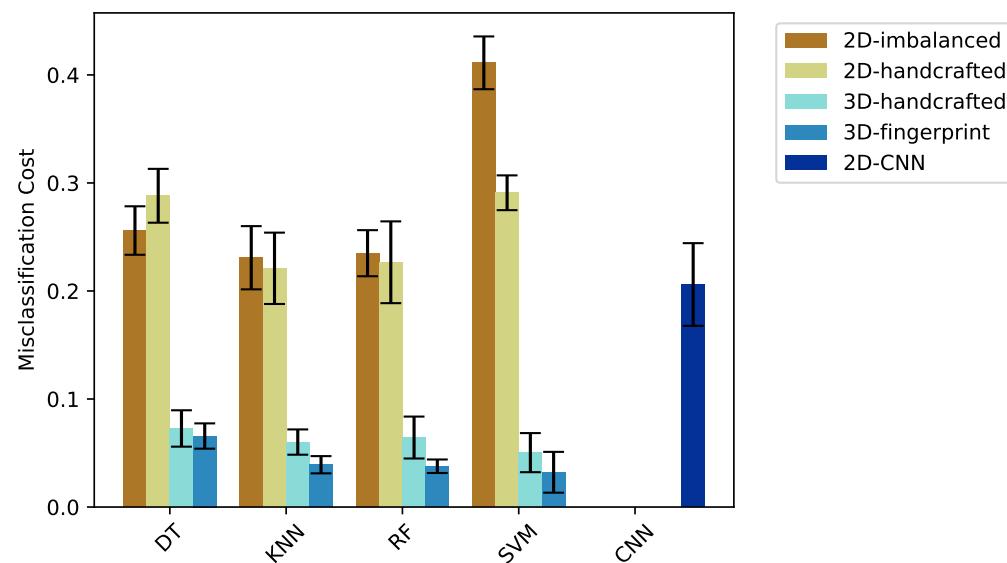
enhanced performance consistency. This analysis underscores the substantial improvement in classification accuracy upon implementing data balancing techniques.

We observed that CNN and SVM fingerprints yielded the most promising results after comparing various models and representations for 2D and 3D datasets. Notably, the 3D data consistently outperformed its 2D counterpart regarding accuracy. The comparison of misclassification costs is presented in Figure 8, and mean per class accuracy is shown in Figure S2. For 2D images, the CNN method is the best-performing classifier with a low misclassification cost compared to other methods. For 3D data, the fingerprint representation and SVM method provide low misclassification costs around eight times lower than the costs of 2D representations. Any 3D representations have less misclassification cost compared to 2D representations. In addition, the confusion matrix in Figure 9 reveals specific areas where 3D data are beneficial, such as misclassifying class 4 as five and class 3 as two.

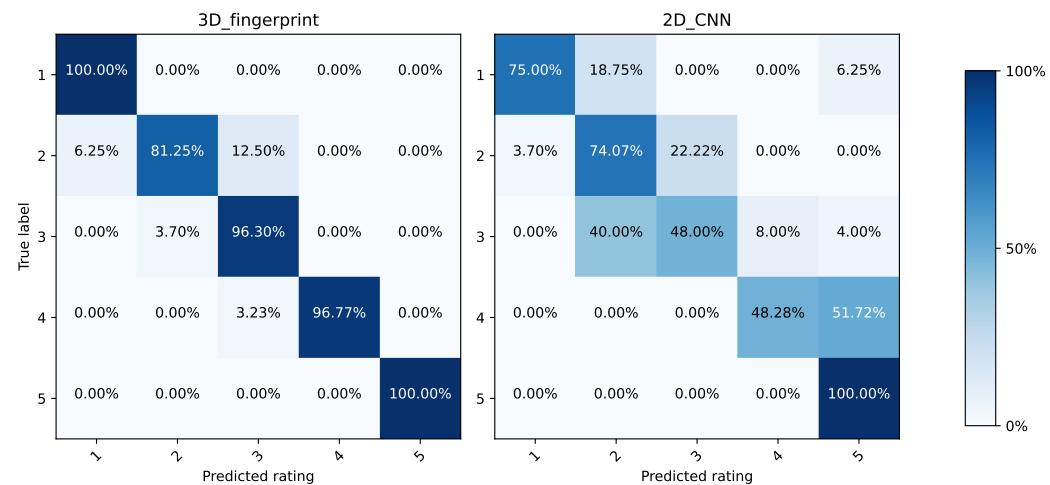
Research has shown that 3D features provide better results than 2D features. For instance, Rios-Toledo et al. [35] found that 3D features outperformed 2D, as the 3D model could recognize plant stress even when visual signals were not apparent. In contrast, a 2D approach only recognizes plant stress in images where the plant shows obvious physical signs of stress. For a plant experiencing first-day stress, the physical signs were not visible in the 2D image, but the morphological decline of the leaves as a result of crop stress was detected by the 3D methodology. This information allowed the 3D methodology to recognize stress even when visual signals were absent on the plant. This performance discrepancy points towards 2D's limitations in comprehensively capturing disease symptoms evident across the canopy.

**Table 3.** Comparing Model Performance on 2D data: Mean Per-Class Accuracy (MPCA) Before and After Data Balancing.

| Model                        | Imbalanced    | Balanced      |
|------------------------------|---------------|---------------|
| Decision Trees (DT)          | 0.749 (0.029) | 0.749 (0.028) |
| K-Nearest Neighbors (KNN)    | 0.780 (0.031) | 0.798 (0.049) |
| Random Forest (RF)           | 0.770 (0.024) | 0.778 (0.028) |
| Support Vector Machine (SVM) | 0.634 (0.034) | 0.707 (0.013) |



**Figure 8.** Comparing Mis-Classification Cost for Different Models on 2D and 3D Data: The evaluated models encompass 2D Imbalanced Features, 2D Handcrafted Color Features, 2D CNN, 3D Handcrafted Color Features and 3D Fingerprints. The x-axis represents the various models (DT, KNN, RF, SVM, and CNN), and the bar plot illustrates the misclassification cost generated by each model.



**Figure 9.** Comparing Confusion Matrices for the Top Two Performing Models Using 2D and 3D Features.

### 3.4. Future Directions

Further enhancements, such as fine-tuning fingerprint representations and incorporating additional canopy information like shape fingerprints, hold promise for improving classifiers. By investigating the correlation between stress and shape fingerprints, the impact of stress on the size and shape of the plant can be analyzed. Also, identifying the most informative sub-canopy in making disease classification can help to develop a hypothesis. While this study evaluated a single time point, a more in-depth and temporal fingerprint could be developed to evaluate the stress progression and the impact of shape and development over time. These temporal fingerprints could provide new insights into agronomic traits, diseases, and pesticide applications, and even develop fingerprint responses to abiotic stress and amendments.

Overall, utilizing 3D canopy data in genomic [55,56] and phenomic research can significantly enrich the comprehension of stress factors. This is especially true in identifying biotic stress traits such as disease and insects [56,57]. Exploring the feature vector within the fingerprint representation, which evolves with plant height, offers the potential for functional GWAS [55] to uncover potential loci associated with multi-scale canopy features linked to stresses. Researchers can efficiently sift through genetic material by employing unique canopy fingerprints, revealing correlations between canopy structures and stress responses. Understanding the impact of various canopy levels on stress factors, such as disease susceptibility and pest infestation, yields valuable insights for refining crop improvement strategies. Leveraging 3D canopy data facilitates more precise prediction of stress responses, thereby enhancing the efficacy of genomic studies to mitigate stress-related agricultural losses. The 3D representation can be stored as metadata to the original point clouds offering a practical solution for querying data and enhancing privacy in deep models, as demonstrated by [58], or stored as a compressed representation of the point cloud that can be used with other modalities to make decision.

While this work primarily focuses on fingerprints assessed from TLS laser point clouds, the concept of canopy fingerprints can be applied across various technologies capable of capturing 3D point clouds like handheld scanners, structured light or constructing 3D point clouds from 2D images like such as space-carving, structure from motion [59–61]. Additionally, advancements in structure-from-motion and related methods, such as Neural Radiance Fields [62] and Gaussian Splatting [63], have made it possible to reconstruct 3D data from a set of 2D multi-view images from 2D images in wild conditions [64].

We note that canopy fingerprints (especially the vertically aligned fingerprints) may produce differing representations if plants are tilted. However, this could be useful for estimating agronomic traits like lodging. Fingerprint representation of stress serves as a

novel strategy for high throughput phenotyping with applications in data curation, cultivar selection, evaluation, and additional experimentation. Integration of canopy fingerprints with machine learning models can further advance the field of phenomics and cyber-agricultural systems [65].

#### 4. Conclusions

This study explores the potential of using 3D point cloud data for precise plant stress evaluation in field crops. We demonstrated that relatively simple (and interpretable) stress representations extracted from 3D point cloud data can be used with classical machine learning models to obtain accurate IDC predictions. In addition, techniques to rectify class imbalance improve classification accuracy and are useful for real-world data complexities where class imbalance is inevitable. The 3D canopy fingerprinting representation with the SVM approach produces accuracies of around 95%, outperforming complicated deep learning models that use only 2D data. This underscores how the integration of 3D data enhances the accuracy of stress identification and classification, as other studies have also shown [66,67]. The findings of this study have important implications for precision agriculture and plant stress evaluation in field crops. The use of 3D point cloud data and classical machine learning models with simple stress representations can lead to more accurate and efficient identification and classification of plant stress, ultimately leading to better crop management and increased yields.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy14061181/s1>, Figure S1: Comparing Handcrafted Features of 2D images: Rating-Specific Distribution Pre- and Post-SVMSMOTE Implementation. Each point illustrates handcrafted color features of IDC-stressed soybean canopies. Colors denote disease ratings, as depicted in Figure 5, highlighting SMOTE's role in amplifying low-instance class samples; Figure S2: Comparison of Classification MPCA for Different Models on 2D and 3D Datasets: The models evaluated include 2D Imbalanced, 2D Handcrafted Color Features, 3D Handcrafted Color Features, 3D Fingerprint, and 2D CNN (utilizing latent features from CNN). The x-axis represents the various models (KNN, RF, SVM, and CNN), and the bar plot illustrates the 5-fold MPCA achieved by each model.

**Author Contributions:** Conceptualization, A.S., B.G. and T.Z.J.; methodology, T.J.Y. and T.Z.J.; software, T.J.Y., S.C. and T.Z.J.; validation, T.J.Y., S.C. and T.Z.J.; formal analysis, T.J.Y. and T.Z.J.; investigation, T.J.Y. and T.Z.J.; resources, A.K.S., A.S. and B.G.; data curation, D.E., T.J.Y. and T.Z.J.; writing—original draft preparation, T.J.Y. and T.Z.J.; writing—review and editing, D.E., A.K.S. and B.G.; visualization, T.J.Y. and T.Z.J.; supervision, B.G.; project administration, B.G.; funding acquisition, A.S., S.S. and B.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project was supported by the Iowa Soybean Association, R.F. Baker Center for Plant Breeding, Plant Sciences Institute, USDA-NIFA Grants#2017-67007-26151, 2017-67021-25965, AI Institute for Resilient Agriculture (USDA-NIFA #2021-67021-35329), COALESCE: COntext Aware LEarning for Sustainable CybEr-Agricultural Systems (NSF CPS Frontier #1954556), FACT: A Scalable Cyber Ecosystem for Acquisition, Curation, and Analysis of Multispectral UAV Image Data (USDA-NIFA #2019-67021-29938), Smart Integrated Farm Network for Rural Agricultural Communities (SIRAC) (NSF S&CC #1952045), and USDA CRIS Project IOW04714.

**Data Availability Statement:** The dataset presented in this study can be found at [https://github.com/znjubery/3D\\_Plant\\_Stress\\_Repo\\_ty](https://github.com/znjubery/3D_Plant_Stress_Repo_ty) (accessed on 24 May 2024).

**Acknowledgments:** The authors thank Soconomics team students. We are thankful to all staff members, particularly Brian Scott and Ryan Dunn, for their assistance with experimentation.

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

## References

1. Hartman, G.L.; West, E.D.; Herman, T.K. Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Secur.* **2011**, *3*, 5–17. [\[CrossRef\]](#)
2. Singh, A.K.; Singh, A.; Sarkar, S.; Ganapathysubramanian, B.; Schapaugh, W.; Miguez, F.E.; Carley, C.N.; Carroll, M.E.; Chiozza, M.V.; Chiteri, K.O.; et al. High-Throughput Phenotyping in Soybean. In *High-Throughput Crop Phenotyping*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 129–163.
3. Fageria, N.K.; Baligar, V.C.; Clark, R.B. Micronutrients in Crop Production. In *Advances in Agronomy*; Sparks, D.L., Ed.; Academic Press: Warsaw, Poland, 2002; Volume 77, pp. 185–268.
4. Froehlich, D.; Fehr, W. Agronomic Performance of Soybeans with Differing Levels of Iron Deficiency Chlorosis on Calcareous Soil 1. *Crop Sci.* **1981**, *21*, 438–441. [\[CrossRef\]](#)
5. Peiffer, G.A.; King, K.E.; Severin, A.J.; May, G.D.; Cianzio, S.R.; Lin, S.F.; Lauter, N.C.; Shoemaker, R.C. Identification of candidate genes underlying an iron efficiency quantitative trait locus in soybean. *Plant Physiol.* **2012**, *158*, 1745–1754. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Castelão Tetila, E.; Brandoli Machado, B.; Belete, N.A.; Guimarães, D.A.; Pistori, H. Identification of Soybean Foliar Diseases Using Unmanned Aerial Vehicle Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2190–2194. [\[CrossRef\]](#)
7. Mahlein, A.K. Plant Disease Detection by Imaging Sensors—Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Dis.* **2016**, *100*, 241–251. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Rumpf, T.; Mahlein, A.K.; Steiner, U.; Oerke, E.C.; Dehne, H.W.; Plümer, L. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Comput. Electron. Agric.* **2010**, *74*, 91–99. [\[CrossRef\]](#)
9. Calderón, R.; Navas-Cortés, J.A.; Lucena, C.; Zarco-Tejada, P.J. High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. *Remote Sens. Environ.* **2013**, *139*, 231–245. [\[CrossRef\]](#)
10. Maimaitijiang, M.; Ghulam, A.; Sidike, P.; Hartling, S.; Maimaitiyiming, M.; Peterson, K.; Shavers, E.; Fishman, J.; Peterson, J.; Kadam, S.; et al. Unmanned Aerial System (UAS)-based phenotyping of soybean using multi-sensor data fusion and extreme learning machine. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 43–58. [\[CrossRef\]](#)
11. Bai, G.; Jenkins, S.; Yuan, W.; Graef, G.L.; Ge, Y. Field-based scoring of soybean iron deficiency chlorosis using RGB imaging and statistical learning. *Front. Plant Sci.* **2018**, *9*, 1002. [\[CrossRef\]](#)
12. Naik, H.S.; Zhang, J.; Lofquist, A.; Assefa, T.; Sarkar, S.; Ackerman, D.; Singh, A.; Singh, A.K.; Ganapathysubramanian, B. A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods* **2017**, *13*, 23. [\[CrossRef\]](#)
13. Hassanijalilian, O.; Igathinathane, C.; Bajwa, S.; Nowatzki, J. Rating Iron Deficiency in Soybean Using Image Processing and Decision-Tree Based Models. *Remote Sens.* **2020**, *12*, 4143. [\[CrossRef\]](#)
14. Dobbels, A.A.; Lorenz, A.J. Soybean iron deficiency chlorosis high-throughput phenotyping using an unmanned aircraft system. *Plant Methods* **2019**, *15*, 97. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Paulus, S. Measuring crops in 3D: Using geometry for plant phenotyping. *Plant Methods* **2019**, *15*, 103. [\[CrossRef\]](#)
16. Jin, S.; Su, Y.; Gao, S.; Wu, F.; Ma, Q.; Xu, K.; Ma, Q.; Hu, T.; Liu, J.; Pang, S.; et al. Separating the structural components of maize for field phenotyping using terrestrial LiDAR data and deep convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 2644–2658. [\[CrossRef\]](#)
17. Jin, S.; Su, Y.; Gao, S.; Wu, F.; Hu, T.; Liu, J.; Li, W.; Wang, D.; Chen, S.; Jiang, Y.; et al. Deep Learning: Individual Maize Segmentation From Terrestrial Lidar Data Using Faster R-CNN and Regional Growth Algorithms. *Front. Plant Sci.* **2018**, *9*, 866. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Zhang, Y.; Yang, Y.; Zhang, Q.; Duan, R.; Liu, J.; Qin, Y.; Wang, X. Toward Multi-Stage Phenotyping of Soybean with Multimodal UAV Sensor Data: A Comparison of Machine Learning Approaches for Leaf Area Index Estimation. *Remote Sens.* **2022**, *15*, 7. [\[CrossRef\]](#)
19. Mahmud, M.S.; He, L. Measuring tree canopy density using A lidar-guided system for precision spraying. In Proceedings of the 2020 ASABE Annual International Virtual Meeting, St. Joseph, MI, USA, 13–15 July 2020.
20. Wijesingha, J.; Moeckel, T.; Hensgen, F.; Wachendorf, M. Evaluation of 3D point cloud-based models for the prediction of grassland biomass. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 352–359. [\[CrossRef\]](#)
21. Zhou, L.; Gu, X.; Cheng, S.; Yang, G.; Shu, M.; Sun, Q. Analysis of plant height changes of lodged maize using UAV-LiDAR data. *Agriculture* **2020**, *10*, 146. [\[CrossRef\]](#)
22. Madec, S.; Baret, F.; de Solan, B.; Thomas, S.; Dutartre, D.; Jezequel, S.; Hemmerlé, M.; Colombeau, G.; Comar, A. High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. *Front. Plant Sci.* **2017**, *8*, 2002. [\[CrossRef\]](#)
23. Zhang, C.; Craine, W.A.; McGee, R.J.; Vandemark, G.J.; Davis, J.B.; Brown, J.; Hulbert, S.H.; Sankaran, S. High-throughput phenotyping of canopy height in cool-season crops using sensing techniques. *Agron. J.* **2021**, *113*, 3269–3280. [\[CrossRef\]](#)
24. Liu, K.; Dong, X.; Qiu, B. Analysis of cotton height spatial variability based on UAV-LiDAR. *Int. J. Precis. Agric. Aviat.* **2020**, *3*, 72–76. [\[CrossRef\]](#)
25. Pagliai, A.; Ammoniaci, M.; Sarri, D.; Lisci, R.; Perria, R.; Vieri, M.; D’Arcangelo, M.E.M.; Storchi, P.; Kartsiotis, S.P. Comparison of Aerial and Ground 3D Point Clouds for Canopy Size Assessment in Precision Viticulture. *Remote Sens.* **2022**, *14*, 1145. [\[CrossRef\]](#)

26. Gu, C.; Zhao, C.; Zou, W.; Yang, S.; Dou, H.; Zhai, C. Innovative Leaf Area Detection Models for Orchard Tree Thick Canopy Based on LiDAR Point Cloud Data. *Agriculture* **2022**, *12*, 1241. [\[CrossRef\]](#)

27. Yun, T.; Cao, L.; An, F.; Chen, B.; Xue, L.; Li, W.; Pincebourde, S.; Smith, M.J.; Eichhorn, M.P. Simulation of multi-platform LiDAR for assessing total leaf area in tree crowns. *Agric. For. Meteorol.* **2019**, *276–277*, 107610. [\[CrossRef\]](#)

28. Su, W.; Zhu, D.; Huang, J.; Guo, H. Estimation of the vertical leaf area profile of corn (*Zea mays*) plants using terrestrial laser scanning (TLS). *Comput. Electron. Agric.* **2018**, *150*, 5–13. [\[CrossRef\]](#)

29. Debnath, S.; Paul, M.; Debnath, T. Applications of LiDAR in Agriculture and Future Research Directions. *J. Imaging* **2023**, *9*, 57. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Rivera, G.; Porras, R.; Florencia, R.; Sánchez-Solís, J.P. LiDAR applications in precision agriculture for cultivating crops: A review of recent advances. *Comput. Electron. Agric.* **2023**, *207*, 107737. [\[CrossRef\]](#)

31. Khanna, R.; Schmid, L.; Walter, A.; Nieto, J.; Siegwart, R.; Liebisch, F. A spatio temporal spectral framework for plant stress phenotyping. *Plant Methods* **2019**, *15*, 13. [\[CrossRef\]](#) [\[PubMed\]](#)

32. Behmann, J.; Mahlein, A.K.; Paulus, S.; Kuhlmann, H.; Oerke, E.C.; Plümer, L. Calibration of hyperspectral close-range pushbroom cameras for plant phenotyping. *ISPRS J. Photogramm. Remote Sens.* **2015**, *106*, 172–182. [\[CrossRef\]](#)

33. Roscher, R.; Behmann, J.; Mahlein, A.K.; Dupuis, J.; Kuhlmann, H.; Plümer, L. Detection of Disease Symptoms on Hyperspectral 3D Plant Models. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *III-7*, 89–96. [\[CrossRef\]](#)

34. Mulugeta Aneley, G.; Haas, M.; Köhl, K. LIDAR-Based Phenotyping for Drought Response and Drought Tolerance in Potato. *Potato Res.* **2023**, *66*, 1225–1256. [\[CrossRef\]](#)

35. Ríos-Toledo, G.; Pérez-Patricio, M.; Cundapí-López, L.Á.; Camas-Anzueto, J.L.; Morales-Navarro, N.A.; Osuna-Coutiño, J.A.d.J. Plant Stress Recognition Using Deep Learning and 3D Reconstruction. In *Pattern Recognition; Lecture Notes in Computer Science*; Rodríguez-González, A.Y., Pérez-Espinosa, H., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Olvera-López, J.A., Eds.; Springer: Cham, Switzerland, 2023; pp. 114–124. [\[CrossRef\]](#)

36. Smith, D.T.; Potgieter, A.B.; Chapman, S.C. Scaling up high-throughput phenotyping for abiotic stress selection in the field. *Theor. Appl. Genet.* **2021**, *134*, 1845–1866. [\[CrossRef\]](#) [\[PubMed\]](#)

37. Minervini, M.; Scharr, H.; Tsafaris, S.A. Image Analysis: The New Bottleneck in Plant Phenotyping [Applications Corner]. *IEEE Signal Process. Mag.* **2015**, *32*, 126–131. [\[CrossRef\]](#)

38. Mitra, N.J.; Guibas, L.J.; Giesen, J.; Pauly, M. Probabilistic Fingerprints for Shapes. In Proceedings of the Symposium on Geometry Processing, Cagliari, Italy, 26–28 June 2006.

39. Probst, D.; Reymond, J.L. A probabilistic molecular fingerprint for big data settings. *J. Cheminform.* **2018**, *10*, 66. [\[CrossRef\]](#) [\[PubMed\]](#)

40. Smith, J.; Smith, O. Fingerprinting Crop Varieties. In *Advances in Agronomy*; Academic Press: Warsaw, Poland, 1992; Volume 47, pp. 85–140. [\[CrossRef\]](#)

41. Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R.P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. *arXiv* **2015**, arXiv:1509.09292.

42. Sadeghi-Tehran, P.; Virlet, N.; Sabermanesh, K.; Hawkesford, M.J. Multi-feature machine learning model for automatic segmentation of green fractional vegetation cover for high-throughput field phenotyping. *Plant Methods* **2017**, *13*, 103. [\[CrossRef\]](#) [\[PubMed\]](#)

43. Young, T.J.; Jubery, T.Z.; Carley, C.N.; Carroll, M.; Sarkar, S.; Singh, A.K.; Singh, A.; Ganapathysubramanian, B. “Canopy fingerprints” for characterizing three-dimensional point cloud data of soybean canopies. *Front. Plant Sci.* **2023**, *14*, 114153. [\[CrossRef\]](#) [\[PubMed\]](#)

44. Faro Scene. 2023. Available online: <https://www.faro.com/en/Products/Software/SCENE-Software> (accessed on 8 October 2023).

45. Zhou, Q.Y.; Park, J.; Koltun, V. Open3D: A modern library for 3D data processing. *arXiv* **2018**, arXiv:1801.09847.

46. Joshi, S.; Jignasu, A.; Young, T.; Elango, D.; Jubery, T.Z.; Jones, S.; Balu, A.; Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; et al. Virtual Reality Assisted Stress Tolerance Rating of Soybean Varieties. In Proceedings of the Fourth International Workshop on Machine Learning for Cyber-Agricultural Systems (MLCAS2022), Iowa State University, Chicago, IL, USA, 10 October 2022.

47. Quinlan, J.R. Improved Use of Continuous Attributes in C4.5. *JAIR* **1996**, *4*, 77–90. [\[CrossRef\]](#)

48. Breiman, L. *Classification and Regression Trees*; Routledge: London, UK, 2017.

49. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)

50. Geva, S.; Sitte, J. Adaptive nearest neighbor pattern classification. *IEEE Trans. Neural Netw.* **1991**, *2*, 318–322. [\[CrossRef\]](#)

51. Hao, P.Y.; Chiang, J.H.; Chen, Y.D. Possibilistic classification by support vector networks. *Neural Netw.* **1995**, *149*, 40–56. [\[CrossRef\]](#) [\[PubMed\]](#)

52. Raschka, S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *arXiv* **2020**, arXiv:1811.12808.

53. Demidova, L.; Klyueva, I. SVM classification: Optimization with the SMOTE algorithm for the class imbalance problem. In Proceedings of the 2017 6th Mediterranean Conference on Embedded Computing (MECO), Bar, Montenegro, 11–15 June 2017; pp. 1–4. [\[CrossRef\]](#)

54. Sun, J.; Li, H.; Fujita, H.; Fu, B.; Ai, W. Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. *Inf. Fusion* **2020**, *54*, 128–144. [\[CrossRef\]](#)

55. Shook, J.M.; Zhang, J.; Jones, S.E.; Singh, A.; Diers, B.W.; Singh, A.K. Meta-GWAS for quantitative trait loci identification in soybean. *G3 Genes Genomes Genet.* **2021**, *11*, jkab117. [\[CrossRef\]](#) [\[PubMed\]](#)

56. Raardin, A.; Fotouhi, F.; Zhang, J.; Mueller, D.S.; Ganapathysubramanian, B.; Singh, A.K.; Dutta, S.; Sarkar, S.; Singh, A. Deep learning-based phenotyping for genome wide association studies of sudden death syndrome in soybean. *Front. Plant Sci.* **2022**, *13*, 966244. [[CrossRef](#)] [[PubMed](#)]
57. Pangga, I.B.; Hanan, J.; Chakraborty, S. Pathogen dynamics in a crop canopy and their evolution under changing climate. *Plant Pathol.* **2011**, *60*, 70–81. [[CrossRef](#)]
58. Cho, M.; Nagasubramanian, K.; Singh, A.K.; Singh, A.; Ganapathysubramanian, B.; Sarkar, S.; Hegde, C. Privacy-preserving deep models for plant stress phenotyping. In Proceedings of the AI for Agriculture and Food Systems, 2022, Vancouver, BC, Canada, 28 February 2021.
59. Nguyen, T.T.; Slaughter, D.C.; Max, N.; Maloof, J.N.; Sinha, N. Structured Light-Based 3D Reconstruction System for Plants. *Sensors* **2015**, *15*, 18587–18612. [[CrossRef](#)] [[PubMed](#)]
60. Das Choudhury, S.; Maturu, S.; Samal, A. Leveraging Image Analysis to Compute 3D Plant Phenotypes Based on Voxel-Grid Plant Reconstruction. *Front. Plant Sci.* **2020**, *11*, 521431. [[CrossRef](#)] [[PubMed](#)]
61. Feng, J.; Saadati, M.; Jubery, T.; Jignasu, A.; Balu, A.; Li, Y.; Attigala, L.; Schnable, P.S.; Sarkar, S.; Ganapathysubramanian, B.; et al. 3D reconstruction of plants using probabilistic voxel carving. *Comput. Electron. Agric.* **2023**, *213*, 108248. [[CrossRef](#)]
62. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106. [[CrossRef](#)]
63. Kerbl, B.; Kopanas, G.; Leimkühler, T.; Drettakis, G. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *arXiv* **2023**, arXiv:2308.04079.
64. Arshad, M.A.; Jubery, T.; Afful, J.; Jignasu, A.; Balu, A.; Ganapathysubramanian, B.; Sarkar, S.; Krishnamurthy, A. Evaluating NeRFs for 3D Plant Geometry Reconstruction in Field Conditions. *arXiv* **2024**, arXiv:2402.10344.
65. Sarkar, S.; Ganapathysubramanian, B.; Singh, A.; Fotouhi, F.; Kar, S.; Nagasubramanian, K.; Chowdhary, G.; Das, S.K.; Kantor, G.; Krishnamurthy, A.; et al. Cyber-agricultural systems for crop breeding and sustainable production. *Trends Plant Sci.* **2024**, *29*, 130–149. [[CrossRef](#)] [[PubMed](#)]
66. Ziamtsov, I.; Navlakha, S. Machine Learning Approaches to Improve Three Basic Plant Phenotyping Tasks Using Three-Dimensional Point Clouds. *Plant Physiol.* **2019**, *181*, 1425–1440. [[CrossRef](#)]
67. Seidel, D.; Annighöfer, P.; Thielman, A.; Seifert, Q.E.; Thauer, J.H.; Glatthorn, J.; Ehbrecht, M.; Kneib, T.; Ammer, C. Predicting Tree Species From 3D Laser Scanning Point Clouds Using Deep Learning. *Front. Plant Sci.* **2021**, *12*, 635440. [[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.