Article

Exploring the Potential of High-Resolution Satellite Imagery for the Detection of Soybean Sudden Death Syndrome

Muhammad M. Raza [1], Chris Harding [2], Matt Liebman [3] and Leonor F. Leandro [1],*

[1]   Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA 50011, USA;
      mraza@iastate.edu
[2]   Department of Geological and Atmospheric Sciences, Iowa State University, Ames, IA 50011, USA;
      charding@iastate.edu
[3]   Department of Agronomy, Iowa State University, Ames, IA 50011, USA; mliebman@iastate.edu
*   Correspondence: lleandro@iastate.edu; Tel.: +1-515-294-8855

check for
updates

Abstract: Sudden death syndrome (SDS) is one of the major yield-limiting soybean diseases in the Midwestern United States. Effective management for SDS requires accurate detection in soybean fields. Since traditional scouting methods are time-consuming, labor-intensive, and often destructive, alternative methods to monitor SDS in large soybean fields are needed. This study explores the potential of using high-resolution (3 m) PlanetScope satellite imagery for detection of SDS using the random forest classification algorithm. Image data from blue, green, red, and near-infrared (NIR) spectral bands, the calculated normalized difference vegetation index (NDVI), and crop rotation information were used to detect healthy and SDS-infected quadrats in a soybean field experiment with different rotation treatments, located in Boone County, Iowa. Datasets collected during the 2016, 2017, and 2018 soybean growing seasons were analyzed. The results indicate that spectral features, when combined with ground-based information, can detect areas in soybean plots that are at risk for disease, even before foliar symptoms develop. The classification of healthy and diseased soybean quadrats was >75% accurate and the area under the receiver operating characteristic curve (AUROC) was >70%. Our results indicate that high-resolution satellite imagery and random forest analyses have the potential to detect SDS in soybean fields, and that this approach may facilitate large-scale monitoring of SDS (and possibly other economically important soybean diseases). It may also be useful for guiding recommendations for site-specific management in current and future seasons.

Keywords: soybean disease; sudden death syndrome; disease detection; remote sensing; PlanetScope; satellite imagery; random forest

---

1. Introduction

Plant disease epidemics cause substantial economic losses in agricultural settings worldwide [1]. Monitoring plant health and detecting plant diseases at early growth stages are essential to control disease spread and facilitate sustainable, environment-friendly, and cost-effective management practices [2] in growers' fields. This paper focuses on the potential of using satellite imagery to detect soybean sudden death syndrome (SDS), a disease of significant economic importance in North and South America [3–5]. The disease is caused by *Fusarium virguliforme* (*Fv*), a soilborne fungal pathogen [6], and is widely distributed across 23 U.S. states, including those accounting for most U.S. soybean production. During the years from 2006 to 2014, annual yield losses in the U.S. due to SDS were estimated at between 0.6 and 1.9 million metric tons [7], representing 200 to 750 million dollars in monetary losses [8].

The pathogen starts infecting roots during early soybean growth stages [9,10] and causes root rot and poor root development [3]. Root infections are favored by cool, wet soil environments [11,12]. Foliar symptoms of interveinal chlorosis (yellowing between leaf veins) and necrosis (tissue browning following cell death) typically appear during reproductive stages [13] and cause premature defoliation and senescence under severe disease pressure [14]. The initial foliar symptoms show only as yellow traces on lower leaves, which makes the disease difficult to detect at early stages. Abundant soil moisture favors SDS foliar symptom expression [12,15], whereas infected plants may not develop foliar symptoms under dry field conditions. Disease distribution within a field is limited by the spatial distribution of the pathogen at the beginning of the growing season.

Scouting for SDS foliar symptoms in the field is made difficult by the relatively late onset of visible foliar symptom expression, which often occurs after the soybean canopy has closed, and by the patchy distribution of SDS in soybean fields. Scouting for symptomatic plants is time-consuming, and confirmation of *Fv* infection requires destructive sampling [16]. Therefore, a more effective method for monitoring and quantifying the distribution of SDS in the field is needed.

Early detection of plant diseases through remote sensing can be difficult when foliar symptoms are mild [16], because multiple factors can contribute to the biophysical and chemical changes that are associated with plant disease [17]. Plants with severe foliar symptoms differ significantly in canopy color from healthy or slightly infected plants [18]. Thus, plants with highly visible symptoms can be more accurately detected than plants with mild symptoms [19].

SDS causes foliar symptoms of chlorosis and necrosis. Canopy damage is the result of disease-induced changes in physiological functions, leaf pigments [19], chlorophyll content, and biomass [16]. Zwiggelaar [20] stated that alteration in leaf pigment content results in changes in optical spectra, such as a decrease in canopy reflectance in the near-infrared band and an increase of reflectance in the red band [21]. Both the red and NIR bands in the light spectrum respond to changes in chlorophyll pigment. These two bands are also highly sensitive to changes in leaf water stress [22] and the total green leaf area [23]. Reduced green leaf area may be present due to reduced plant growth during mid-season or due to early defoliation at the end of the season.

Leaf chlorophyll level is inversely related to spectral reflectance in the blue and red wavebands of visible light, and is dependent on the interactions between plant tissues and absorbed light [24]. SDS-infected plants tend to show early chlorophyll deterioration [16], resulting in spectral responses that increase reflectance in the blue and red bands while decreasing reflectance in the green and NIR regions. When plant diseases such as SDS produce necrotic or chlorotic symptoms on leaves, diseased plants show an overall increase in reflectance in the visible region, especially in the red band [16,25], and simultaneously decreased reflectance in the green and NIR regions. Reflectance in the NIR regions is affected primarily by leaf structure and canopy biomass, which are influenced by plant water and health status [22,25–27].

Recently, successful early detection of SDS in soybean leaves [16,25] and canopies [25,28] has been achieved using handheld, tractor-mounted, and unmanned aerial vehicle (UAV)-mounted remote sensing tools. Although these tools have proven to be promising for early detection of SDS, they still require considerable time, capital, and equipment investments for data collection. Also, these cutting-edge tools can be expensive and cumbersome for growers to operate and maintain in a commercial system. Data generated from these tools may lack spatial information, which must be obtained via additional geographic information system (GIS) and global positioning system (GPS) technologies before targeted and site-specific management applications can be implemented.

Satellite imagery, in comparison with ground-based and aerial remote sensing platforms, covers a wide ground swath and can provide high temporal resolution due to frequent revisit times. Consequently, there has been a growing interest in using high-resolution multispectral satellite images to monitor crop diseases [29–31]. Image acquisition costs vary considerably depending on the provider and product specifications, but more low-cost or free image products are available now than in the past. For example, PlanetScope is a 4-band multispectral satellite that collects imagery at

3-m spatial resolution with 1- to 10-day temporal resolution (Planet Labs https://www.planet.com/). PlanetScope ortho scenes are geo-rectified (i.e., processed to remove distortions caused by tilt and terrain), radiometrically and atmospherically corrected, and projected.

Different machine learning classification methods, such as support vector machines (SVM), artificial neural networks (ANNs), linear discriminant analysis (LDA), and random forest (RF) [16,25,29,32–35], have been used for early detection of plant diseases based on remote sensing data. Random forest is a flexible and powerful machine learning classifier [36] that has been utilized in the classification of remote-sensing-based information [29,37–41]. The random forest classifier can handle huge, multidimensional datasets and performs both classification and regression functions without over-fitting the model [36,42]. The random forest algorithm also evaluates the predictive importance of input features, such as reflectance wavebands in our study, hence supporting feature selection for subsequent analysis [36].

Greater access to low-cost, multispectral, high-resolution satellite imagery provides opportunities for scientists to combine these large datasets with machine learning technologies [27] to devise reliable automated systems for detecting and mapping plant diseases. These systems could support grower decisions for site-specific management applications, such as planting resistant soybean varieties and using diverse crop rotations [43] in infested areas. They have the potential to significantly reduce harmful chemical seed treatment applications and the resulting economic expense and ecological impact on soybean production systems.

The objective of this study was to detect SDS using high-resolution satellite imagery collected over different vegetative and reproductive growth stages of the soybean crop in a long-term crop rotation field trial, with a history of rotation-influenced differences in SDS symptom intensity [44,45]. We hypothesized that spectral data based on blue, green, red, near-infrared (NIR) spectral bands, and the calculated normalized difference vegetation index (NDVI), combined with ground-based crop management information, would allow detection of SDS in larger field-scale plots. Specifically, we wished to investigate the potential of using:

(1)   high-resolution satellite images to assess soybean crop health at the canopy scale, which would permit inspection of larger field areas.
(2)   RF algorithms to classify disease status using image data from single dates collected over a two-month period, to see if and when subtle changes in canopy reflectance properties not associated with typical foliar symptoms would permit classification of crop disease status.
(3)   knowledge of field areas likely to exhibit differences in disease intensity.

2. Materials and Methods

We used 4-band PlanetScope multispectral imagery for the 2016 to 2018 soybean growing seasons, along with crop rotation information, and analyzed the data using the random forest algorithm for SDS detection. Our method for data processing and analysis for the detection SDS is illustrated in Figure 1. The method encompasses six key components: (1) image acquisition and data extraction, (2) normalized difference vegetation index (NDVI) calculation, (3) exploration and preparation of input datasets, (4) training and inspection of the random forest model, (5) model classification of diseased or healthy status, and (6) assessment of classification accuracy.
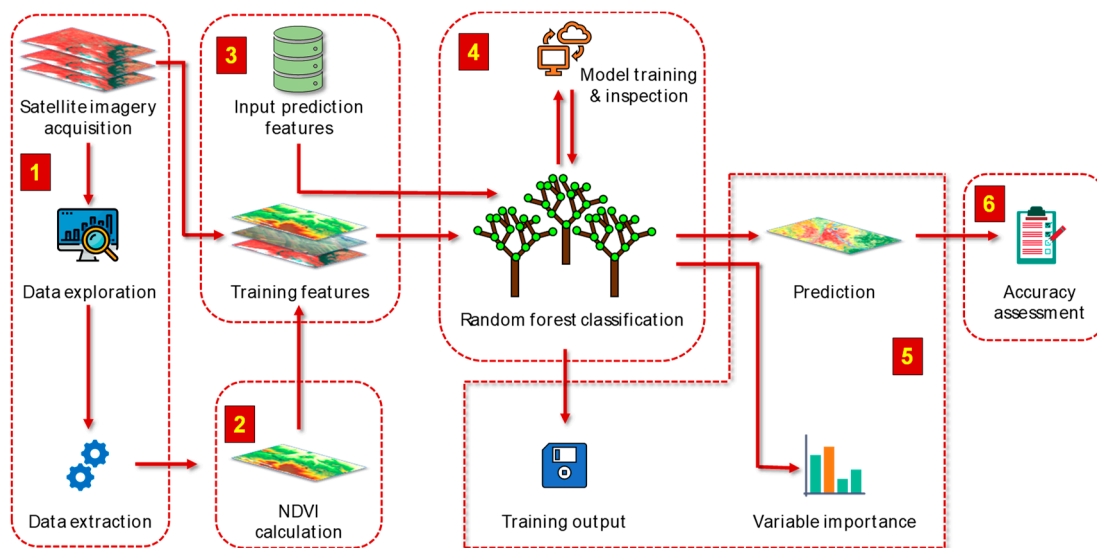
Figure 1. Workflow of the automated data processing and analysis steps employed to perform forest-based classification. This methodology was divided into six main steps for a better understanding of this tool's workability and to facilitate future studies. Note: normalized difference vegetation index—NDVI.

2.1. Study Site

Data for this study were collected from an ongoing soybean field experiment located at Iowa State University's Marsden Farm in Boone County, Iowa [44,45]. This experiment was established in 2001 as a randomized complete block design (RCBD) with four blocks. Each block consists of nine main plots (18 m × 84 m), with two subplots (9 m × 84 m) within each main plot (Figure 2).

Three crop rotation systems (2-year, 3-year, and 4-year) were applied to the main plots [43]. The 2-year cropping system was planted with corn and soybean in alternate years and managed with extension-recommended practices and applications of synthetic fertilizers. In the 3-year cropping system, an oat-red clover forage mix was planted in the year after soybean and before corn. In the 4-year cropping system, oat and alfalfa were seeded in the 3rd year and followed by alfalfa in the 4th year. Crop nutrients for the 3-year and 4-year cropping systems were provided via applications of composted cattle manure and reduced rates of synthetic fertilizers. All crop phases of the three rotation systems were planted each year.

No artificial *Fv* inoculum was added to the field plots in this experiment. However, in every year since 2010, the crop rotation treatments have been associated with differences in SDS intensity at this site [15]. Therefore, this site was chosen to test using satellite imagery for SDS detection.

The soybean variety Latham L2758 R2, a maturity group 2.7 variety rated as resistant to SDS, was planted on 21 May in 2016, 15 May in 2017, and 17 May in 2018. Each soybean plot consisted of 24 rows, with row spacing of 0.76 m. Plots were subdivided into 20 quadrats, each roughly 8 × 9 m (Figure 2). For visual disease assessment, rating areas (3 m wide × 1.5 m long) were flagged inside each quadrat area, 8.4 m apart in the outer six rows along both sides of the plot, resulting in 20 rating areas in each soybean main plot. Center coordinates for each quadrat were collected using a handheld Trimble GeoXT GPS Device (Trimble Inc., Sunnyvale, CA, USA). Each visual rating area spanned four soybean rows.
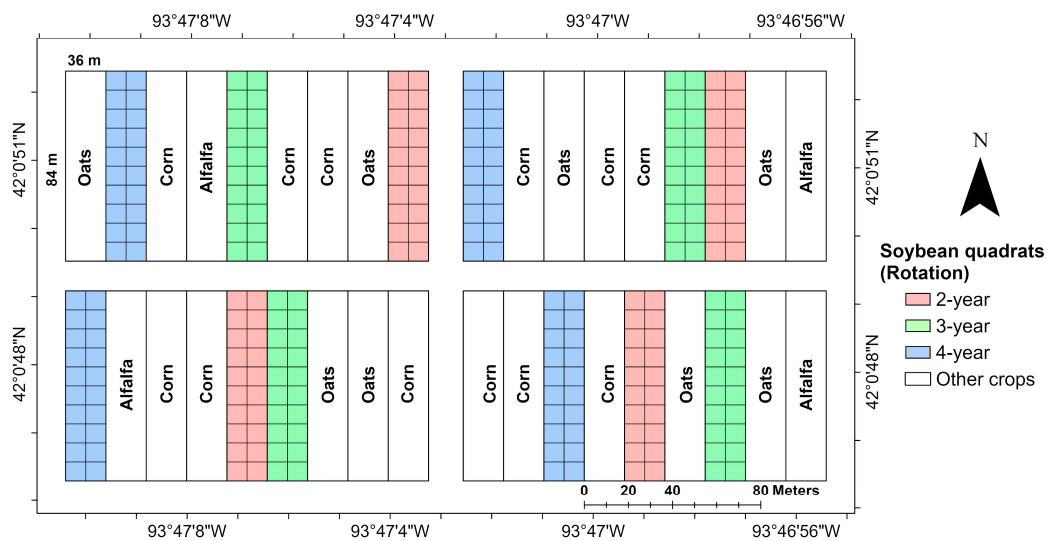
Figure 2. Overview of the Marsden Farm study site, located in Boone County, Iowa, in 2016. The randomized complete block design had four blocks with nine main plots, which were subdivided into two subplots. Subplots were divided into 10 quadrats (roughly 8 × 9 m, shown as grid squares). Main plots were planted with soybean in 2-year, 3-year, and 4-year crop rotations in 2016. Other plots were planted with corn, alfalfa, oat, and red clover. Soybean plot locations were prescribed by crop rotation sequences, which were assigned to main plots.

2.2. Disease Assessment

SDS foliar incidence, i.e., the percentage of plants in a quadrat with foliar symptoms [46], was assessed weekly in all soybean quadrats, from the time of initial foliar symptom onset in the trial until crop senescence (Figure 3). For analysis, soybean quadrats were assigned to two main classes, i.e., healthy or diseased, based on foliar symptom incidence. Soybean quadrats with no foliar symptoms or less than 5% foliar disease incidence were classified as healthy quadrats, while quadrats showing more than 5% foliar incidence at the end of the season were classified as diseased quadrats. Disease was rated by a team of plant pathologists, including researchers who had been studying SDS at this site since 2010. All plants within the visual rating areas (1.5 m L × 3 m W) were examined. Disease was confirmed in the field by examining a few plants for internal stem discoloration (which distinguishes foliar symptoms caused by SDS from symptoms that may be caused by the brown stem rot pathogen, *Phialophora gregata*) [3], and by molecular detection of the pathogen in soybean roots that were collected for concurrent studies [42] (L. Leandro, unpublished data). Plants were dug using hand trowels to minimize soil disturbance and loss of adjacent plants. Foliar symptoms were confirmed when disease symptoms in quadrats were noted for two or more consecutive rating dates. Plant stand data for the same quadrats were used to convert counts of diseased plants to percentages.

2.3. Imagery Acquisition

PlanetScope (PS) satellite imagery was acquired from Planet Labs Inc. (San Francisco, CA, USA), a private imaging company. PS satellites orbit at 400 km (51.6° inclination) altitude and can provide images with red, green, and blue (RGB) and near-infrared (NIR) data (Table 1). The image product is delivered as a continuous, split-frame strip, with half-frames containing RGB and NIR imagery.
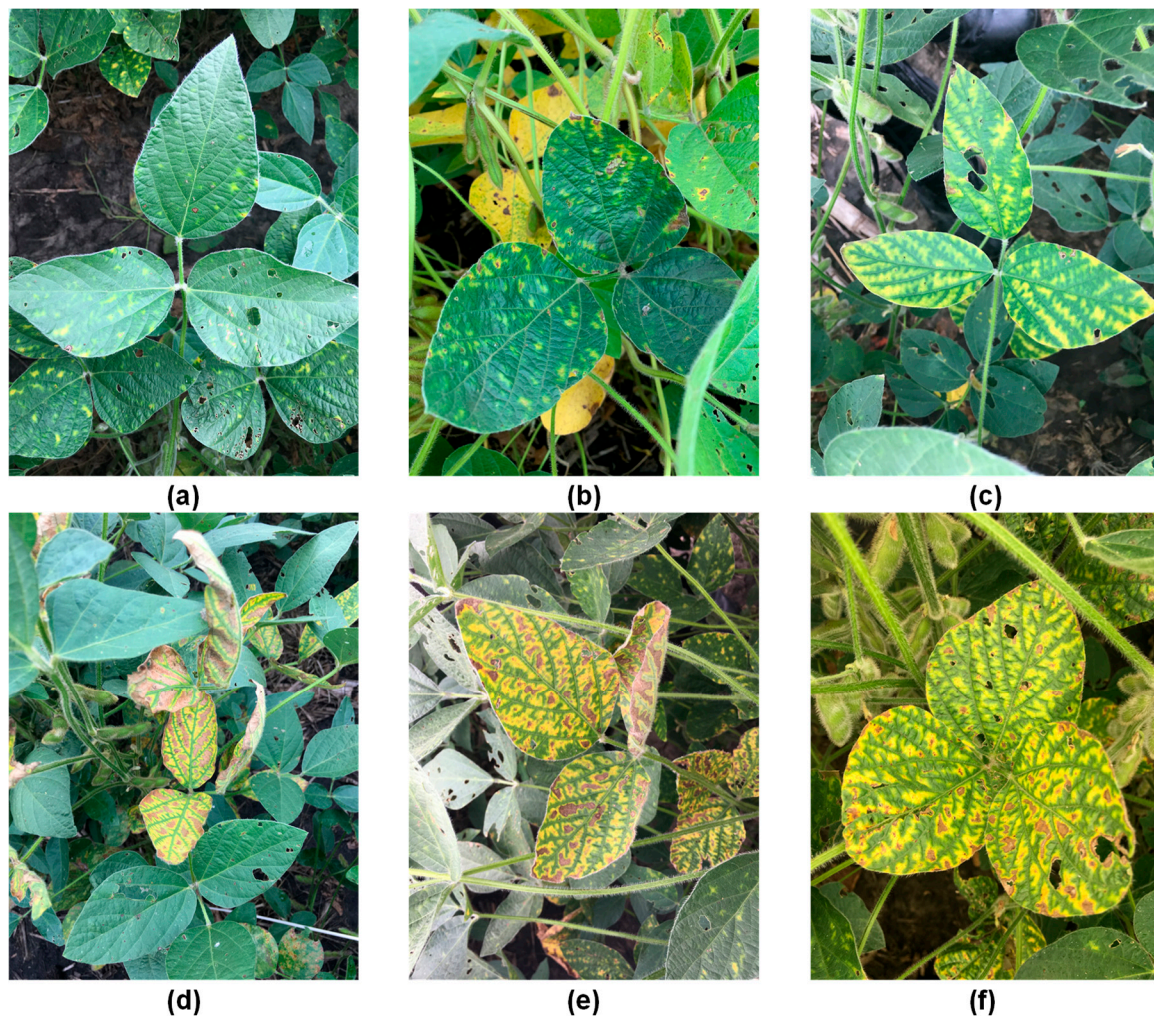
Figure 3. Foliar symptoms of soybean sudden death syndrome. Foliar symptom development starts as scattered yellow spots (**a**,**b**) which later enlarge (**c**) and turn into large regions of interveinal chlorosis and necrosis on leaves (**d**–**f**).

Table 1. Description of spectral bands of PlanetScope satellite ortho scenes used in this study.

PlanetScope Band	Spectrum Region	Wavelength (nm)	Spatial Resolution
Band_1	Blue	455–515	
Band_2	Green	500–590	3 m
Band_3	Red	590–670	
Band_4	NIR	780–860	

We utilized PS ortho scenes (Planet Labs image product Level 3B), which are orthorectified (i.e., processed to remove geometric distortions caused by tilt and terrain) and radiometrically and sensor corrected. The geometric correction was based on digital elevation models (DEMs) with post-spacing between 30 and 90 m, with an error specification of <10 Root Mean Square Error (RMSE). Ortho scenes were projected to the universal transverse mercator (UTM) WGS84 cartographic projection prior to product delivery. The PS ortho tile product included a single strip of multiple orthorectified scenes that were merged and divided according to a defined grid. The 3-m resolution ortho scenes were obtained for the months of July and August in all years. We obtained 4, 15, and 18 cloud-free ortho scenes in 2016, 2017, and 2018, respectively. All 2016 images were created by generation 0e (A) satellites and all 2018 images by generation 0f or 10 (B) satellites. A and B have different spectral

response curves. Beginning with 20-Jul-2017, the 2017 images switched from A to B, with the sole exception of 23-Aug-2017, which again used A.

Image alignment was checked visually in ArcGIS and compared to ground-based GIS reference points. Differences in image brightness and color—due to differences in satellite sensor hardware, acquisition time, sun angle, cloud conditions, and dates—are likely to be present, and were noted in the images we used. Because our intent was to evaluate a method for analyzing data from single dates, we did not standardize image brightness during pre-processing.

2.4. Data Extraction

Soybean plot locations varied from year to year and followed crop rotation sequences assigned to main plots. Quadrat locations for each year were described by shapefiles, which included crop rotation information (2-year, 3-year, or 4-year) and quadrat health status at the end of the season (SDS incidence reached 5% or higher, or plants remained healthy). For data extraction from satellite images and subsequent analysis, soybean quadrats were generalized to large quadrats (8.6 m W × 9.1 m L) (Figure 2). Via the ArcPy Python Application Programming Interface (API) [47], the ArcGIS Zonal statistics tool was used to calculate the mean values of all image pixels that were covered by each quadrat; mean digital number (DN) values were calculated for each of the four bands (R, G, B, and NIR). This process was repeated for each of the images obtained for a given year, so that each quadrat polygon feature contained the crop rotation type, the mean R, G, B, and NIR values of the pixels covered by the polygon, a calculated NDVI value (explained below), and a binary value for SDS health status.

Preliminary inspection of the image data indicated that band 4 (NIR) mean DN values for the two disease classes were significantly different in 24 of 37 images, as determined by Wilcoxon's rank test (two-tailed test, $p \leq 0.05$) performed via PROC NPAR1WAY i.e., the procedure to perform Wilcoxon's rank test in Statistical analysis system (SAS Institute, Cary, NC, USA). Similar tests for bands 1, 2, and 3 indicated 13, 10, and 17 images showed class differences when quadrat mean DN values were tested. Exploratory analyses detected class differences in more of the images when single pixel data were examined (data not shown). Crop rotation, R, G, B, NIR, and NDVI were then used as exploratory variables (corresponding to "features" in the random forest methodology), and SDS status was used as the "ground truth" response variable.

A practical example of the data extraction process (zonal statistics) via Python is given in the Jupyter Notebook "Prepare data for random forest classification.ipynb", which can be found in the data repository: https://doi.org/10.25380/iastate.11356430 [48]. The repository also contains gis project data in a ESRI (Environmental Systems Research Institute) file geodatabase (.gdb); specifically a set of preprocessed quadrat polygons and images (cropped to cover just the area under investigation) for a selection of dates in 2016, 2017, and 2018. Finally, the repository contains an ArcGIS Pro project file (.aprx) with which the data can be visualized. See supplementary materials for more information.

2.5. NDVI Calculation

Although several conventional and newly-modified vegetation indices based on the red-edge bands have been explored to evaluate crop health [16,25,29,32], NDVI [49] is one of the most commonly used vegetation indices. To evaluate its suitability for detecting and discriminating healthy and diseased soybean quadrats, the NDVI was calculated for each quadrat using Equation (1):

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \tag{1}$$

2.6. Data Analysis

Random forest classification [36] was performed using R and the Scikit-learn [50] library in Python 3.7.3 [51]. Random forest is a supervised machine learning algorithm that applies bootstrap aggregation (bagging) [52,53] and random feature selection [54,55]. Random forest is a robust, nonparametric

method that builds a forest of many decision trees from a training dataset and tests the accuracy of the forest on a test dataset. Although individual decision trees may not be accurate, the combination of many trees significantly increases the robustness of the model and leads to more accurate predictions. When the test datasets are passed through the trained decision trees, each decision tree votes on an outcome. The forest model then aggregates the votes from all decision trees and predicts the final class based on the largest number of votes. In our case, the outcome is in the form of a binary variable (healthy, SDS = 0, or diseased, SDS = 1); thus, we performed a random forest classification [36] rather than a regression.

2.6.1. Random Forest Model Training and Prediction

Data from each ortho scene were analyzed separately, beginning by splitting the data into training and test datasets (70% and 30% of the quadrats, respectively), followed by model training and classification of the test data. Model parameters included the number of trees in the forest (n_estimators), number of features to consider when looking for the best split (max_features), the maximum tree depth allowed (max_depth), and minimum number of samples required to be a leaf node (min_sample_leaf).

These parameters were tuned using the 5-fold cross-validation grid search method. Grid search is a simple strategy used to evaluate all possible combinations of given, discrete parameter spaces and derive their optimal combination for a tuned model. Using these tuned models, SDS predictions were made using the test set. This procedure was repeated for each ortho scene in 2016, 2017, and 2018.

The data repository https://doi.org/10.25380/iastate.11356430 contains a Jupyter Notebook "Using Random Forest Models for SDS.ipynb", which demonstrates the entire process of splitting the data into training and test datasets, tuning the random forest model on the training dataset, performing the predictions of the test dataset using the tuned model, and judging the prediction quality of the model on the quadrat data discussed earlier.

2.6.2. Variable Importance

We measured the importance of predictor variables via the permutation variable importance measure procedure [56], which is based on random selection and index reordering. First, the random forest model calculates prediction accuracy in the out-of-bag (OOB) observations (observations from the training set that are left out of the bootstrap samples and not used to construct the decision trees), and then it randomly shuffles values of a predictor variable to break the association between response and predictor values and recalculate the accuracy in OOB observations. Then, it calculates the difference in model accuracy before and after shuffling [57]. The average of this number over all trees in the forest represents the raw importance score for the variable. If the predictor never had any meaningful relationship with the response, shuffling its values will produce very little change in the model accuracy. However, if a predictor was strongly associated with the response, permutations should create a significant decrease in accuracy. Permutation of variable importance evaluates the individual impact, including interactions, of each tree predictor in the random forest model. The procedure was used to evaluate all variables.

2.6.3. Accuracy Assessment

Confusion matrices (Table 2) were constructed to assess the precision, specificity, sensitivity, and overall accuracy [58] of SDS predictions made based on data from each PS ortho scene. Precision describes the percentage of "diseased" classifications that were assigned correctly. Specificity explains the percentage of correctly classified healthy quadrats, and sensitivity describes the percentage of correctly classified diseased quadrats. The overall accuracy is given by the proportion of correctly classified quadrats (of either status), which explains the ability of random forest trained models to classify healthy and diseased quadrats correctly. In addition, the area under the receiver operating characteristic (AUROC) curve was computed to evaluate the overall model performance for predicting SDS in the unknown test set [59]. The receiver operating curve is a graphical plot that

summarizes the performance of a classifier (i.e., a model) over all possible thresholds, which indicates the diagnostic ability of the model. This process was repeated for all ortho scenes collected from the three soybean growing seasons.

Table 2. Example of a confusion matrix, based on data for an ortho scene collected on July 3, 2017, computed to evaluate the performance of random forest trained models in detecting healthy and sudden death syndrome infected soybean quadrats. Disease "negative" and "positive" classes refer to quadrats that are healthy or diseased, respectively.

Parameters		Ground Truth Data		
		Healthy	**Diseased**	
Prediction	Healthy	True negatives (30)	False negatives (6)	
	Diseased	False positives (5)	True positives (31)	Precision (86%)
Recall		Specificity (86%)	Sensitivity (84%)	Accuracy (85%)

3. Results

3.1. Disease Onset and Progress

SDS foliar symptom onset was first seen at 66 (July 27), 94 (August 17), and 92 (August 17) days after planting (DAP) in 2016, 2017, and 2018, respectively (Figure 4). In all seasons, foliar SDS symptoms were first observed at reproductive soybean growth stages. Environmental conditions were more conducive for SDS in the 2016 growing season, with earlier onset of foliar symptoms and greater foliar incidence than in 2017 and 2018. SDS foliar symptom development started late in the cropping season in 2017 and 2018 (Figure 4). Early season rainfall was much higher than normal during 2018, which resulted in prolonged flooding in the south end of two of the soybean plots in block 4. Due to flooding damage, data from 12 quadrats were omitted from analysis of the 240 quadrats in the trial in 2018.
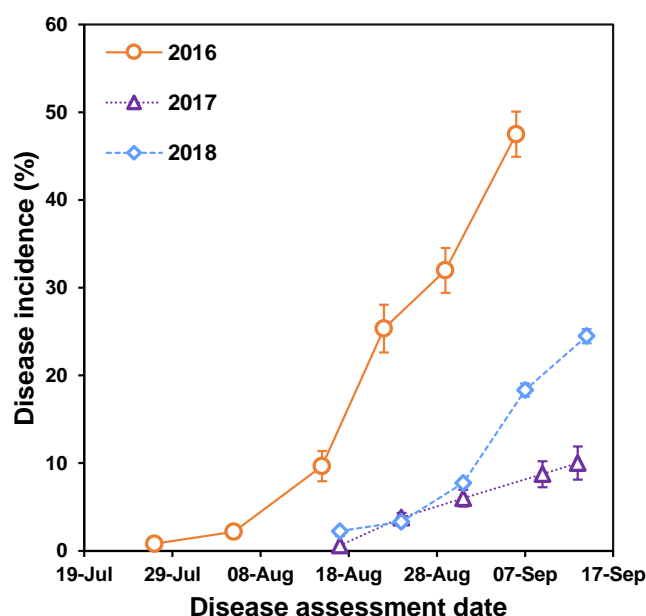

Figure 4. Progress of soybean sudden death syndrome (SDS) foliar disease incidence (%) over time in the 2016, 2017, and 2018 soybean growing seasons at Marsden Farm. Each dot represents the overall mean incidence of diseased plants in quadrats classified as diseased (i.e., quadrats where >5% of the plants had foliar symptoms) on a given date, calculated from 240 (2016, 2017) or 228 (2018) quadrats in soybean plots. Error bars indicate standard error of mean disease incidence.

In two of the three years, SDS incidence progressed slowly for the first two weeks after symptom onset and then increased suddenly (Figure 4). Foliar symptoms were first seen as slight chlorotic mottling on lower leaves, which was followed by rugosity and marginal cupping of upper canopy leaves. As foliar damage progressed, leaves developed interveinal chlorosis and necrosis. Severely affected plants lost leaflets prematurely, leaving petioles attached. Frequencies of healthy and diseased quadrats observed at the end of each soybean growing season are shown in Figure 5. In 2016 and 2017, 2-year soybean rotation plots had the highest frequencies of diseased quadrats at the end of the season (53% and 56%, respectively), and there were fewer diseased quadrats in 3-year and 4-year rotation plots. At the end of the 2018 growing season, the percentage of quadrats classified as diseased was high in all rotation treatments (Figure 5).
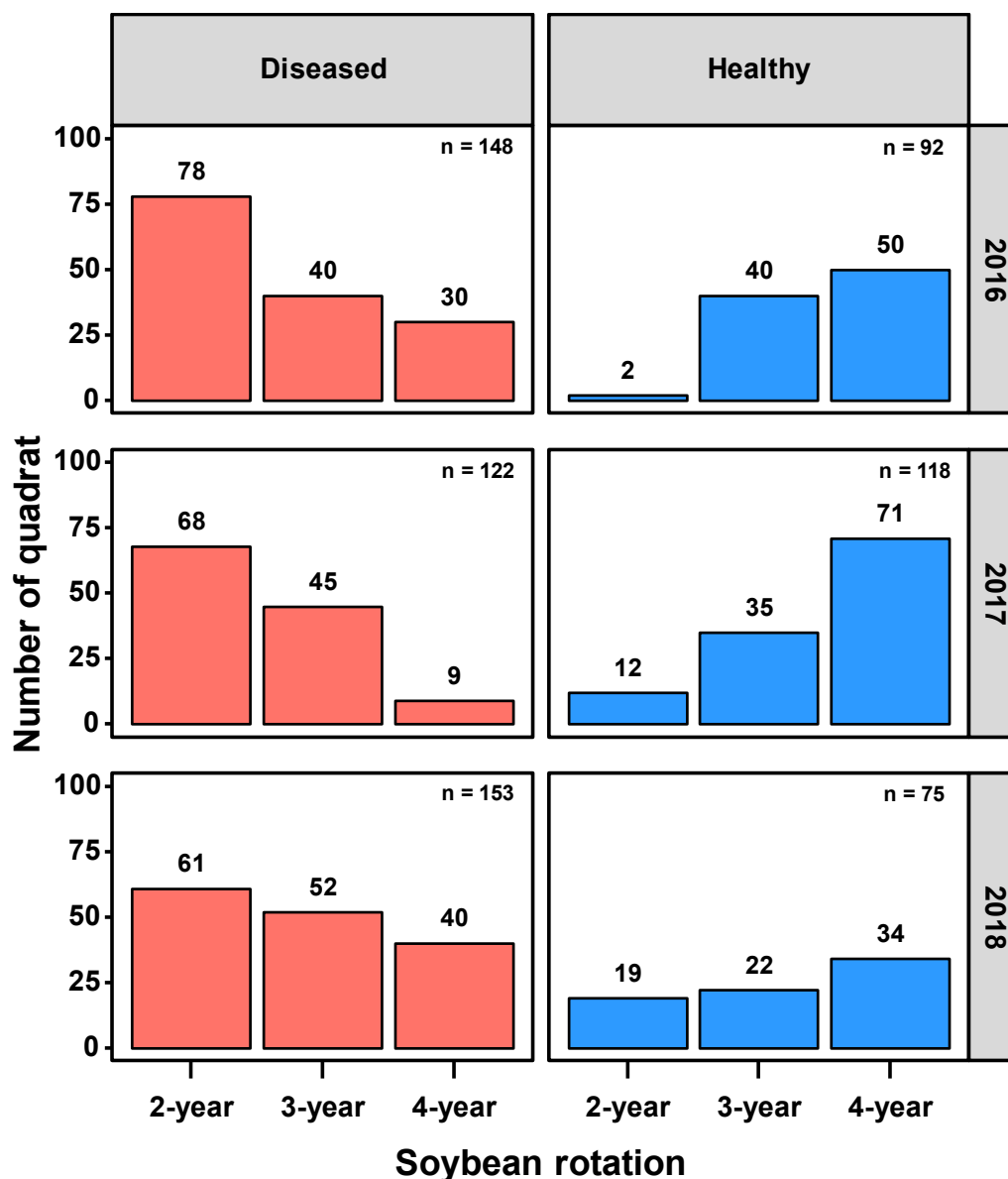

Figure 5. Frequencies of healthy and diseased quadrats in 2-, 3-, and 4-year soybean rotation plots at the end of the 2016 to 2018 soybean growing seasons. Quadrats with no foliar SDS symptoms and less than 5% foliar incidence were classified as healthy quadrats, and quadrats with foliar symptoms on more than 5% of the plants at the end of the season were classified as diseased quadrats. In 2018, 12 flooded soybean quadrats were omitted from data analysis.

3.2. Modeling SDS with Random Forest

3.2.1. Predictive Importance of Input Variables

The relative predictive importance of all explanatory variables included in the model for detecting healthy and diseased soybean quadrats is shown in Figure 6. Crop rotation information was important for discriminating between diseased and healthy soybean quadrats in the 2016 and 2017 growing seasons, but less important in 2018, due to similar frequencies of diseased and healthy quadrats in the three rotation treatments (Figure 5).
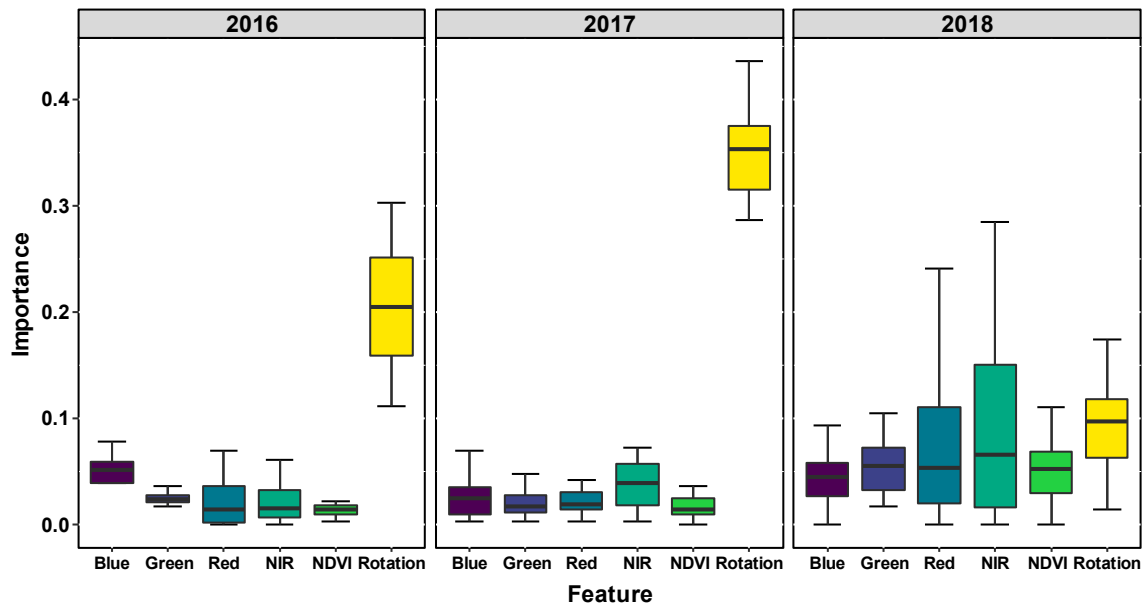
Figure 6. Distribution of predictive importance of features (B, G, R, and NIR image data from PlanetScope ortho scenes; NDVI; and crop rotation information) used to detect healthy and SDS-infected soybean quadrats in 2016, 2017, and 2018 crop growing seasons.

Despite the comparatively high predictive importance of the crop rotation feature, it alone does not lead to accurate classification results. A direct comparison of results of RF models using all features, with spectral features alone (without crop rotation information), and with crop rotation alone (without spectral features) suggests that the combination of spectral features with crop rotation information provides the most accurate models. The results of these comparisons are shown in the Appendix A.

3.2.2. Precision

Random forest-trained models classified diseased soybean quadrats with high precision (i.e., >0.77) in all ortho scenes collected in the three soybean growing seasons (Figure 7a). Diseased quadrats were consistently classified in all years, with an overall precision of 0.77–0.97 (Table 3).

3.2.3. Sensitivity and Specificity

Random forest-trained models provided consistent, high classification specificity (0.58–0.96) and sensitivity (0.70–0.95) in all soybean growing seasons (Table 3). In the majority of ortho scenes, the sensitivity of SDS detection was higher than the specificity (Figure 7b). These statistics indicate that the diseased quadrats were more accurately classified than the healthy quadrats in the majority of the ortho scenes. However, in some scenes, sensitivity remained lower than specificity (Figure 7b), which means that healthy quadrats were more accurately classified than diseased quadrats in those ortho scenes.
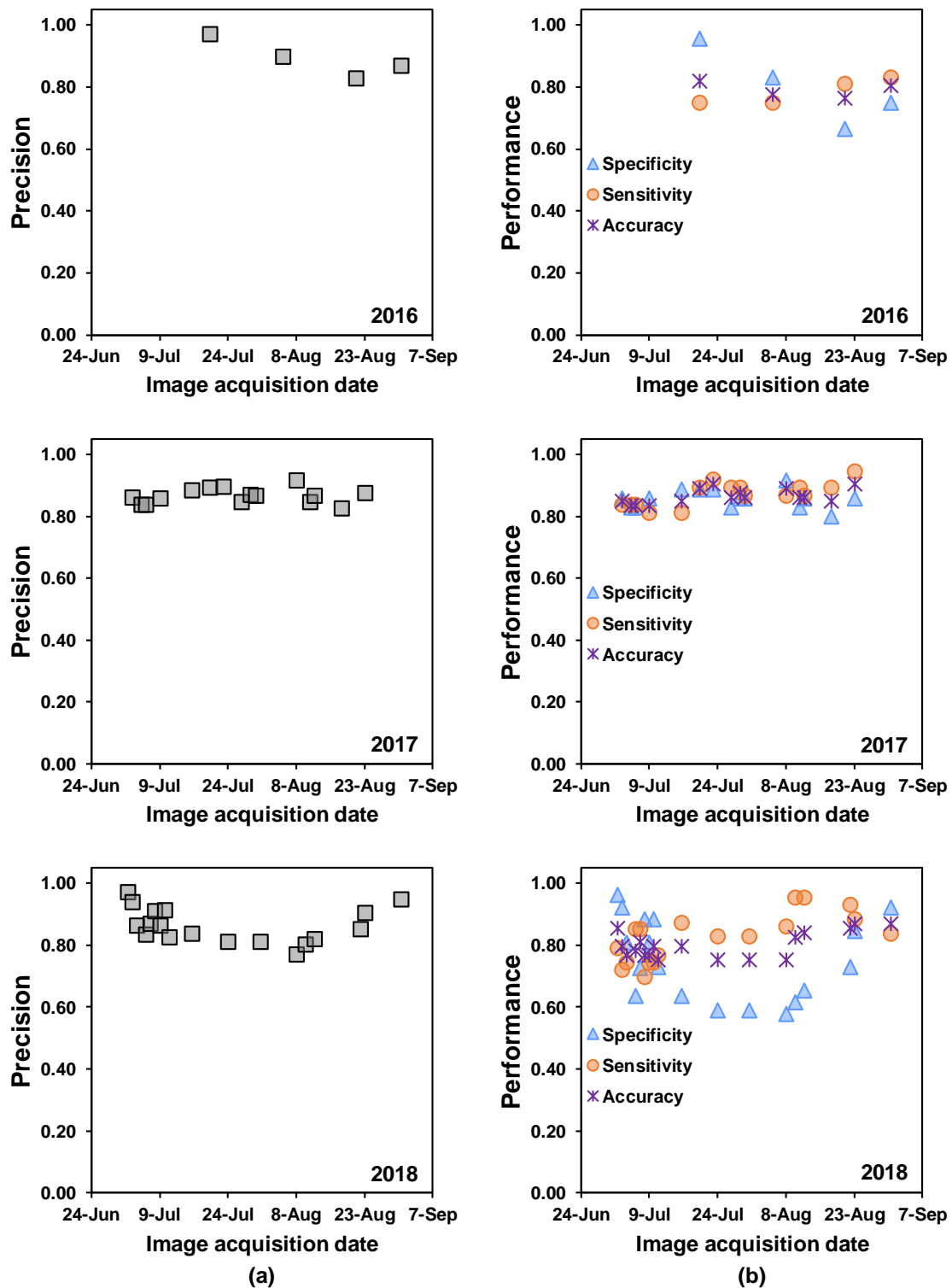
Figure 7. (**a**) Classification precision (proportion of quadrats classified as "diseased" that were assigned correctly), (**b**) specificity (proportion of correctly classified healthy quadrats), sensitivity (proportion of correctly classified diseased quadrats), and accuracy (proportion of observations classified correctly) achieved by random forest trained models for SDS detection in soybean quadrats in 2016 through 2018. In all growing seasons, SDS-infected quadrats were detected with high accuracy, even before the onset of visible foliar symptoms.

Table 3. Classification precision, specificity, sensitivity, overall accuracy, and area under receiver operating characteristic curve (AUROC) achieved in classifying healthy and SDS-infected soybean quadrats. This table shows the range, i.e., the minimum and maximum values obtained for these parameters in all years.

Year	Precision	Specificity	Sensitivity	Accuracy	AUROC
2016	0.83–0.97	0.67–0.96	0.75–0.83	0.76–0.82	0.81–0.90
2017	0.83–0.91	0.80–0.91	0.81–0.95	0.83–0.90	0.87–0.94
2018	0.77–0.97	0.58–0.96	0.70–0.95	0.75–0.87	0.70–0.89

3.2.4. Classification Accuracy

Quadrat health status (healthy or diseased) was classified correctly in at least 75% of the quadrats on all dates (Table 3) during the 3-year study period. Classification accuracy for individual dates ranged from 75 to 90%. High classification accuracy was achieved using data from ortho scenes that were collected before foliar symptoms were visible. The results slightly varied over the years, but overall remained consistent over the season (Figure 7).

3.2.5. Area under Receiver Operating Characteristic Curve (AUROC)

Overall, AUROC values ranged from 0.70 to 0.94 (Table 3). High AUROC values were obtained using satellite images from the first week of July in 2017 and 2018 growing seasons, before the onset of visible SDS foliar symptoms (Figure 8a). In 2016 and 2017, AUROC values for quadrat disease classification were consistently high, i.e., >80% (Figure 8b). In 2018, AUROC values were lowest for classification of data obtained between July 16 and August 8, before the onset of visible symptoms, but AUROC values for the 2018 growing season ranged from 0.70 to 0.89 (Table 3).
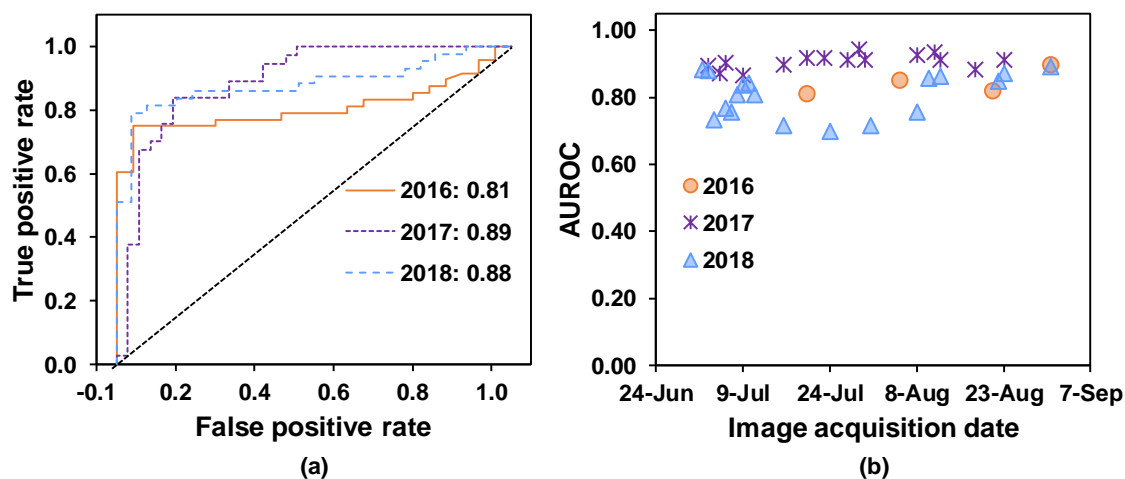

Figure 8. (**a**) Examples of receiver operating characteristic (ROC) plots for classification of SDS-infected soybean quadrats based on reflectance data from satellite images obtained in 2016 (July 20), 2017 (July 3), and 2018 (July 2). Values after colons indicate AUROC values. (**b**) AUROC values achieved by random forest trained models used to detect healthy and SDS-infected soybean quadrats.

4. Discussion

Remote sensing data have the potential to detect SDS-stress-triggered changes in the pattern of light emission from plant canopies [60]. Our findings suggest that canopy changes due to soybean SDS can be detected using high-resolution satellite imagery, even before the development of foliar symptoms in soybean fields (Figure 7), and that prediction accuracy improves when known site-specific factors are included in the model.

Both healthy and diseased soybean quadrats were classified with >77% precision based on PlanetScope 3-m-resolution images collected from 2016 through 2018 i.e., quadrats were correctly classified as diseased or healthy more than 77% of the time. Overall, SDS classification sensitivity was higher than specificity, which indicates that classification accuracy was higher for diseased soybean quadrats than for healthy quadrats.

In exploratory data analyses, we saw that diseased soybean quadrats had high reflectance in the visible light bands (RGB) and low reflectance in the NIR band. However, healthy soybean quadrats showed lower reflectance in the visible light bands and higher reflectance in the NIR band. These patterns became more evident for disease ratings near the end of the season, when SDS foliar symptom onset and incidence increased and loss of leaf biomass and green pigment were readily observed. Despite the low level of incidence and foliar symptoms in the 2017 and 2018 growing seasons, trained random forest models detected diseased quadrats with high accuracy, even when plants were at vegetative growth stages.

In all three years, SDS-symptomatic soybean quadrats were classified correctly with 75% or better classification accuracy (Table 3). Moreover, in 2016 and 2017, diseased and healthy soybean quadrats were detected with more than 81% AUROC, indicating that random forest trained models detected SDS with good discriminative ability, even before foliar symptom development. An AUROC value of 1.0 represents perfect classification, whereas an AUROC value of 0.5 indicates pure chance.

ROC curves are useful tools for measuring classification accuracy in binary-class problems such as ours [61]. In 2018, AUROC values decreased for satellite images obtained between July 16 and August 8 and improved in subsequent images. Foliar damage caused by application of post-emergent herbicides (glyphosate as potassium salt (1.54 kg active ingredient/ha) plus lactofen (0.14 kg active ingredient/ha)) on July 11, 2018, at soybean growth stage R1-R2, likely altered soybean canopy reflectance during this period. Plant stress and foliar tissue damage were observed after the spray, and we postulate that this could have interfered with SDS detection in images collected during that time period. Despite these facts, the overall AUROC values ranged from 0.70 to 0.89, demonstrating the random forest models' ability to detect SDS-associated changes with reasonable accuracy.

The random forest algorithm allows the evaluation of the predictive importance of the different input features (exploratory variables). By testing different feature combinations for SDS detection, we found high and consistent accuracy for AUROC values when using all features to train the models. The importance of features (variables) varied. In many cases, crop rotation information was more important than any of the spectral features. However, this alone was never sufficient for good results. For the best results, a set of all features (i.e., spectral and known environmental or treatment factors) should be used. It should be left to the RF model to prioritize certain features from this set to arrive at an optimal outcome.

A recent study conducted at the same site suggests that the long-term, diversified crop rotation systems have a significant effect on SDS foliar severity and incidence [43]. Because rotation-specific differences in SDS incidence have been documented since 2010, we included rotation information in the SDS detection model along with remote sensing variables. The importance of crop rotation information for SDS detection in our study suggests that features specific to a study site (e.g., prior knowledge of disease occurrence and intensity, landscape features, environmental inputs, or cultivar susceptibility) may be as important as spectral features. Yang et al. [62] used a combination of geographic features and satellite image data to predict SDS risk in commercial soybean fields throughout the state of Iowa. Their field survey data confirmed that information from early-season NDVI values and geographical features (e.g., field slope and distance from water flow lines) were correlated with late-season SDS intensity.

In another study, Herrmann et al. [25] detected SDS before visual symptom onset with a classification accuracy of 88% and 91% from canopy and leaf-scale reflectance spectra, respectively. These classification accuracies were achieved for observations made after mid-July. In comparison, we observed similar or higher levels of SDS detection accuracy for the first week of July in the 2017 and

2018 cropping seasons. Our study differed from that of Herrmann et al. [25] in that they discriminated between artificially inoculated and non-inoculated plots, with unknown levels of naturally occurring inoculum, whereas we assessed naturally infested plots and classified "true" disease status at the end of the growing season.

In another microplot study, Hatton et al. [28] used thermal infrared sensors mounted on an unmanned aerial vehicle to assess SDS development in soybean plots. They conducted flights once foliar symptom development started and found a strong correlation between canopy temperatures and SDS symptom development at the end of the growing season. However, they did not address the early detection of SDS before foliar symptom onset.

5. Limitations and Future Work

Although we obtained high classification accuracies, our study has limitations. Our site already had a multiyear disease history documented by very detailed and diligent ground-truthing. Rotation turned out to be an important predictor at our site, as we have multiple years of data to document differences in SDS disease intensity. However, in other field situations, no scouting may have been performed to establish SDS at specific locations or no history of SDS occurrence may be available.

As we only used the test subset for our prediction, it is not known how useful a model trained on data from prior years would be for predicting disease in fields with unknown or poorly described prior disease histories. Even with a well-documented dataset like ours, additional studies are needed to evaluate if a model trained with data from past years could be used to predict SDS in future years.

Another limitation is that we tested only one sampling design for our study, and it is highly likely that sampling strategies could be optimized. Because of the high degree of manual work involved, we ground-truthed only the center (1.5 × 3 m) of each 9 × 9 m quadrat. However, in the random forest model, we used the average of the entire 9 × 9 m quadrat as spectral values for the quadrat. Estimation of disease status in large areas using scouting data may be improved by estimating disease at multiple points within an area of interest, or by utilizing data from areas with a range of disease intensities. Evaluation of sampling designs should be conducted using multiple sites and years, with data collected at both small and large spatial scales.

The current work showed that a consistent separation between the two classes (healthy vs. SDS) in all bands cannot realistically be expected when using satellite images with moderately-large pixels collected on fairly uniform, green plant canopies. The random forest models were able to incorporate small differences in just a few bands (or even one band), but these typically were less important than the crop rotation. However, a greater degree of consistent separation between multiple bands would certainly be desirable. Focusing only on the pixels covering the ground-truthed 1.5 × 3 m area within each quadrat may offer better results in the future. The use of multiband imagery acquired by drones (UAVs), which are becoming more common and provide a far higher resolution, is an alternative we are currently exploring.

Random forest models are known to cope well with a high number of variables (features). Similarly to previous work in this field, our study used NDVI as a variable for the model. However, it is unclear if this is permissible. One might argue that as NDVI is a ratio involving only the red and NIR bands, and as the mean of each of these bands is already used, no new information is added to the model when using NDVI. Instead, it might be more advantageous to explore the inclusion of other variables that describe each of the spectral variables in more detail, such as the minimum, maximum, standard deviation, median, or other quantiles. Our study also did not encode the location of the quadrats relative to each other for the random forest model, and this spatial (distance) information could also be used. Depending on their scale, future studies might potentially include soil types, distance to flowlines [62], and climatic and environmental trends [15].

A minor limitation arises from the satellite generations being switched (with different spectral response curves) in late July 2017. This means that the graphs showing our results along timelines may exhibit a break at that date. However, the impact of this change in response curves on the random forest

models is unclear. While it is true that a model from 2016 is, therefore, not comparable to a 2018 model, we never compared models and their results directly, other than via the afore-mentioned graphs.

In the current work, each RF model was trained on a single date. In the future, we plan to explore the use of models that are based on multiple dates (e.g., all late growing season dates), which we expect should increase the predictive power of RF models. In this scenario, the generation of the satellite will be used as an independent variable when generating the random forest model.

6. Conclusions

In this study, we explored the potential of using multispectral PlanetScope satellite imagery for the classification of soybean quadrats as either healthy or SDS-infected using random forest models with spectral data (blue, green, red, and NIR bands), calculated NDVI, and crop rotation information obtained for three years (2016–2018). We provided example data and Python code to enable others to test, extend, and further develop this technique. Splitting the data into training (70%) and test (30%) subsets, healthy and diseased soybean quadrats were detected with more than 75% accuracy and 70% AUROC in all years, suggesting that high-resolution satellite imagery has the potential for early and accurate detection of SDS in soybean fields.

High-resolution multispectral imagery of cropland represents a relatively untapped, vast resource to explore. Early detection of SDS may be used by the industry to develop novel, sustainable, and environment-friendly means to mitigate this disease at early stages, and may meet practical demands for detection of SDS at a regional scale to inform field management applications and crop insurance decisions. Future studies could evaluate high-resolution multispectral imagery at the regional scale for early detection and monitoring of other economically significant diseases, which may facilitate effective policymaking, support decision making, and guide recommendations for sustainable and efficient allocation of management resources in precision agriculture systems.

Appendix A

The Effect of Features on Accuracy, Precision, Specificity, and Sensitivity

Table A1 compares the results of RF models that use different sets of features (variables) for three dates late in the growing season in 2016, 2017, and 2018: (a) all features (R, B, G, NIR, NDVI, rotation); (b), only spectral data (R, G, B, NIR, NDVI), and (c) crop rotation alone. Each model was separately optimized for its set of features. Accuracy, precision, specificity, and sensitivity were highest for set

a (all features). Removing the rotation (b) resulted in equal or lower values (with a few exceptions). Using only rotation (c) produced the worst results.

Precision and specificity values on the rotation-only model for 20160823 are misleading artifacts of very a lopsided confusion matrix—no healthy quadrats were actually classified correctly and a large part of the SDS quadrats were misclassified as healthy. Similarly, for 20180823, specificity and sensitivity are artifacts—this model classified all quadrats as SDS (i.e., no healthy quadrats were correctly classified).

Table A1. Comparison of random forest (RF) models trained using different sets of features for three dates late in the growing season in 2016, 2017, and 2018.

Date	Features	Accuracy	Precision	Specificity	Sensitivity	Important Features*
20160823						
(a)	B, G, R, NIR, NDVI, Rotation	0.81	0.89	0.79	0.83	Rotation, B, NDVI
(b)	B, G, R, NIR, NDVI	0.76	0.83	0.67	0.81	B, NDVI, G
(c)	Rotation	0.73	**1**	**1**	0.6	Rotation
20170823						
(a)	B, G, R, NIR, NDVI, Rotation	0.83	0.79	0.79	0.68	Rotation, NIR, G
(b)	B, G, R, NIR, NDVI	0.55	0.57	0.5	0.61	NIR, G, R
(c)	Rotation	0.75	0.71	0.59	0.89	Rotation
20180823						
(a)	B, G, R, NIR, NDVI, Rotation	0.77	0.79	0.5	0.89	NDVI, NIR, Rotation
(b)	B, G, R, NIR, NDVI	0.78	0.81	0.55	0.85	NDVI, G, R
(c)	Rotation	0.68	0.68	**0**	**1**	Rotation

Note: * the three most important features in decreasing order of importance. Keys used: G=green; B=blue; R=red; NIR=near-infrared; NDVI=normalized difference vegetation index.

References

1. Madden, L.V.; Hughes, G.; Van den Bosch, F. *The Study of Plant Disease Epidemics*; American Phytopathological Society Press: Saint Paul, MN, USA, 2007; pp. 1–10.
2. Martinelli, F.; Scalenghe, R.; Davino, S.; Panno, S.; Scuderi, G.; Ruisi, P.; Villa, P.; Stroppiana, D.; Boschetti, M.; Goulart, L.R. Advanced methods of plant disease detection. A review. *Agron. Sustain. Dev.* **2015**, *35*, 1–25. [CrossRef]
3. Hartman, G.L.; Leandro, L.F.; Rupe, J.C. Sudden death syndrome. In *Compendium of Soybean Diseases and Pests*, 5th ed.; Hartman, G.L., Rupe, J.C., Sikora, E.J., Domier, L.L., Davis, J.A., Steffey, K.L., Eds.; American Phytopathological Society: Saint Paul, MN, USA, 2015; pp. 88–90.
4. Wrather, A.; Shannon, G.; Balardin, R.; Carregal, L.; Escobar, R.; Gupta, G.K.; Ma, Z.; Morel, W.; Ploper, D.; Tenuta, A. Effect of diseases on soybean yield in the top eight producing countries in 2006. *Plant Health Prog.* **2010**, *10*, 29. [CrossRef]
5. Bandara, A.Y.; Weerasooriya, D.K.; Bradley, C.A.; Allen, T.W.; Esker, P.D. Dissecting the economic impact of soybean diseases in the United States over two decades. *bioRxiv* **2019**. [CrossRef] [PubMed]
6. Aoki, T.; O'Donnell, K.; Homma, Y.; Lattanzi, A.R. Sudden-death syndrome of soybean is caused by two morphologically and phylogenetically distinct species within the *Fusarium solani* species complex—*F. virguliforme* in North America and *F. tucumaniae* in South America. *Mycologia* **2003**, *95*, 660–684. [CrossRef]
7. Allen, T.W.; Bradley, C.A.; Sisson, A.J.; Byamukama, E.; Chilvers, M.I.; Coker, C.M.; Collins, A.A.; Damicone, J.P.; Dorrance, A.E.; Dufault, N.S. Soybean yield loss estimates due to diseases in the United States and Ontario, Canada, from 2010 to 2014. *Plant Health Prog.* **2017**, *18*, 19–27. [CrossRef]
8. Navi, S.S.; Yang, X.B. Sudden death syndrome—A growing threat of losses in soybeans. *CAB Rev.* **2016**, *11*, 1–13. [CrossRef]
9. Gao, X.; Hartman, G.; Niblack, T. Early infection of soybean roots by *Fusarium solani* f. sp. *glycines*. *Phytopathology* **2006**, *96*, S38.
10. Rupe, J.C.; Gbur, E.E. Effect of plant age, maturity group, and the environment on disease progress of sudden death syndrome of soybean. *Plant Dis.* **1995**, *79*, 139–143. [CrossRef]

11. Gongora-Canul, C.C.; Leandro, L.F.S. Effect of soil temperature and plant age at time of inoculation on progress of root rot and foliar symptoms of soybean sudden death syndrome. *Plant Dis.* **2011**, *95*, 436–440. [CrossRef]

12. Scherm, H.; Yang, X.B. Development of sudden death syndrome of soybean in relation to soil temperature and soil water matric potential. *Phytopathology* **1996**, *86*, 642–649. [CrossRef]

13. Navi, S.S.; Yang, X.B. Foliar symptom expression in association with early infection and xylem colonization by *Fusarium virguliforme* (formerly *F. solani* f. sp. *glycines*), the causal agent of soybean sudden death syndrome. *Plant Health Prog.* **2008**, *10*. [CrossRef]

14. Roy, K.W.; Rupe, J.C.; Hershman, D.E.; Abney, T.S. Sudden death syndrome of soybean. *Plant Dis.* **1997**, *81*, 1100–1111. [CrossRef] [PubMed]

15. Leandro, L.F.S.; Robertson, A.E.; Mueller, D.S.; Yang, X.B. Climatic and environmental trends observed during epidemic and non-epidemic years of soybean sudden death syndrome in Iowa. *Plant Health Prog.* **2013**, *14*, 18. [CrossRef]

16. Bajwa, S.G.; Rupe, J.C.; Mason, J. Soybean disease monitoring with leaf reflectance. *Remote Sens.* **2017**, *9*, 127. [CrossRef]

17. Hatfield, J.L.; Gitelson, A.A.; Schepers, J.S.; Walthall, C.L. Application of spectral remote sensing for agronomic decisions. *Agron. J.* **2008**, *100*, 117–131. [CrossRef]

18. Mee, C.Y.; Balasundram, S.K.; Hanif, A.H.M. Detecting and monitoring plant nutrient stress using remote sensing approaches: A review. *Asian J. Plant. Sci.* **2017**, *16*, 1–8.

19. Sankaran, S.; Mishra, A.; Ehsani, R.; Davis, C. A review of advanced techniques for detecting plant diseases. *Comput. Electron. Agric.* **2010**, *72*, 1–13. [CrossRef]

20. Zwiggelaar, R. A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop. Prot.* **1998**, *17*, 189–206. [CrossRef]

21. Jacquemoud, S.; Ustin, S.L. Leaf optical properties: A state of the art. In Proceedings of the 8th International Symposium of Physical Measurements & Signatures in Remote Sensing, Aussois, France, 8–12 January 2001; pp. 223–332.

22. Nilsson, H.E. Remote sensing and image analysis in plant pathology. *Can. J. Plant Pathol.* **1995**, *17*, 154–166. [CrossRef]

23. Mulla, D.J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* **2013**, *114*, 358–371. [CrossRef]

24. Lillesand, T.; Kiefer, R.W.; Chipman, J. *Remote Sensing and Image Interpretation*, 7th ed.; Wiley: New York City, NY, USA, 2014; p. 768.

25. Herrmann, I.; Vosberg, S.K.; Ravindran, P.; Singh, A.; Chang, H.X.; Chilvers, M.I.; Conley, S.P.; Townsend, P.A. Leaf and canopy level detection of *Fusarium virguliforme* (sudden death syndrome) in soybean. *Remote Sens.* **2018**, *10*, 426. [CrossRef]

26. Gates, D.M.; Keegan, H.J.; Schleter, J.C.; Weidner, V.R. Spectral properties of plants. *Appl. Opt.* **1965**, *4*, 11–20. [CrossRef]

27. 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]

28. Hatton, N.; Sharda, A.; Schapaugh, W.; Van der Merwe, D. Remote thermal infrared imaging for rapid screening of sudden death syndrome in soybean. In Proceedings of the 2018 ASABE Annual International Meeting, Detroit, MI, USA, 29 July–1 August 2018.

29. Zheng, Q.; Huang, W.; Cui, X.; Shi, Y.; Liu, L. New spectral index for detecting wheat yellow rust using Sentinel-2 multispectral imagery. *Sensors* **2018**, *18*, 868. [CrossRef]

30. Yuan, L.; Zhang, J.; Shi, Y.; Nie, C.; Wei, L.; Wang, J. Damage mapping of powdery mildew in winter wheat with high-resolution satellite image. *Remote Sens.* **2014**, *6*, 3611–3623. [CrossRef]

31. Yuan, L.; Pu, R.; Zhang, J.; Wang, J.; Yang, H. Using high spatial resolution satellite imagery for mapping powdery mildew at a regional scale. *Precis. Agric.* **2016**, *17*, 332–348. [CrossRef]

32. 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]

33. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef]

34. Rahman, H.U.; Ch, N.J.; Manzoor, S.; Najeeb, F.; Siddique, M.Y.; Khan, R.A. A comparative analysis of machine learning approaches for plant disease identification. *Adv. Life Sci.* **2017**, *4*, 120–126.

35. Wang, X.; Zhang, M.; Zhu, J.; Geng, S. Spectral prediction of *Phytophthora infestans* infection on tomatoes using artificial neural network (ANN). *Int. J. Remote Sens.* **2008**, *29*, 1693–1706. [CrossRef]

36. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

37. Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Dedieu, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* **2016**, *187*, 156–168. [CrossRef]

38. Tian, S.; Zhang, X.; Tian, J.; Sun, Q. Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. *Remote Sens.* **2016**, *8*, 954. [CrossRef]

39. Mahdianpari, M.; Salehi, B.; Mohammadimanesh, F.; Motagh, M. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. *ISPRS J. Photogramm.* **2017**, *130*, 13–31. [CrossRef]

40. Melville, B.; Lucieer, A.; Aryal, J. Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. *Int. J. Appl. Earth Obs.* **2018**, *66*, 46–55. [CrossRef]

41. Liu, Y.; Gong, W.; Hu, X.; Gong, J. Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. *Remote Sens.* **2018**, *10*, 946. [CrossRef]

42. Dietterich, T.G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Mach. Learn.* **2000**, *40*, 139–157. [CrossRef]

43. Leandro, L.F.S.; Eggenberger, S.; Chen, C.; Williams, J.; Beattie, G.A.; Liebman, M. Cropping system diversification reduces severity and incidence of soybean sudden death syndrome caused by *Fusarium virguliforme*. *Plant Dis.* **2018**, *102*, 1748–1758. [CrossRef]

44. Hunt, N.D.; Hill, J.D.; Liebman, M. Reducing Freshwater Toxicity while Maintaining Weed Control, Profits, And Productivity: Effects of Increased Crop Rotation Diversity and Reduced Herbicide Usage. *Environ. Sci. Technol.* **2017**, *51*, 1707–1717. [CrossRef]

45. Hunt, N.D.; Hill, J.D.; Liebman, M. Cropping System Diversity Effects on Nutrient Discharge, Soil Erosion, and Agronomic Performance. *Environ. Sci. Technol.* **2019**, *53*, 1344–1352. [CrossRef]

46. Nutter, F.W., Jr.; Teng, P.S.; Shokes, F.M. Disease assessment terms and concepts. *Plant Dis.* **1991**, *75*, 1187–1188.

47. Toms, S. *ArcPy and ArcGIS–Geospatial Analysis with Python*; Packt Publishing Ltd.: Birmingham, UK, 2015.

48. Harding, C.; Raza, M.M. GIS Data and Juptyer Notebook for Random Forest Models for Soybean Sudden Death Syndrome (SDS). Dataset. Iowa State University, 2019. Available online: https://doi.org/10.25380/iastate.11356430.v1 (accessed on 1 February 2020).

49. Rouse, J.W., Jr.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, Washington, DC, USA, 10–14 December 1973; Nasa Special Publication NASA SP-351. pp. 309–317.

50. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

51. Rossum, G.V.; Drake, F.L. *The Python Language Reference Manual*; Network Theory Ltd.: Bristol, UK, 2011; p. 15.

52. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]

53. Breiman, L. Heuristics of instability and stabilization in model selection. *Ann. Stat.* **1996**, *24*, 2350–2383. [CrossRef]

54. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; pp. 278–282.

55. Barandiaran, I. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.

56. Janitza, S.; Strobl, C.; Boulesteix, A.L. An AUC-based permutation variable importance measure for random forests. *BMC Bioinform.* **2013**, *14*, 119. [CrossRef]

57. Strobl, C.; Malley, J.; Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* **2009**, *14*, 323. [CrossRef]

58. Ting, K.M. Confusion Matrix. In *Encyclopedia of Machine Learning and Data Mining*, 2nd ed.; Sammut, C., Webb, G.I., Eds.; Springer: New York City, NY, USA, 2017; p. 260.

59. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [CrossRef]

60. Chaerle, L.; Van Der Straeten, D. Imaging techniques and the early detection of plant stress. *Trends Plant Sci.* **2000**, *5*, 495–501. [CrossRef]

61. Ben-David, A. About the relationship between ROC curves and Cohen's kappa. *Eng. Appl. Artif. Intell.* **2008**, *21*, 874–882. [CrossRef]

62. Yang, S.; Li, X.; Chen, C.; Kyveryga, P.; Yang, X.B. Assessing field-specific risk of soybean sudden death syndrome using satellite imagery in Iowa. *Phytopathology* **2016**, *106*, 842–853. [CrossRef] [PubMed]