



Article

Hyperspectral Imaging Combined with Machine Learning for the Detection of Fusiform Rust Disease Incidence in Loblolly Pine Seedlings

Piyush Pandey ^{1,2} , Kitt G. Payn ², Yuzhen Lu ³ , Austin J. Heine ² , Trevor D. Walker ² , Juan J. Acosta ² and Sierra Young ^{1,*}

- ¹ Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, NC 27695, USA; ppandey4@ncsu.edu
- ² Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA; kgpayn@ncsu.edu (K.G.P.); ajheine@ncsu.edu (A.J.H.); Trevor_Walker@ncsu.edu (T.D.W.); jjacosta@ncsu.edu (J.J.A.)
- ³ Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, MS 39762, USA; yzlu@abe.msstate.edu
- * Correspondence: syoung2@ncsu.edu

Abstract: Loblolly pine is an economically important timber species in the United States, with almost 1 billion seedlings produced annually. The most significant disease affecting this species is fusiform rust, caused by *Cronartium quercuum* f. sp. *fusiforme*. Testing for disease resistance in the greenhouse involves artificial inoculation of seedlings followed by visual inspection for disease incidence. An automated, high-throughput phenotyping method could improve both the efficiency and accuracy of the disease screening process. This study investigates the use of hyperspectral imaging for the detection of diseased seedlings. A nursery trial comprising families with known in-field rust resistance data was conducted, and the seedlings were artificially inoculated with fungal spores. Hyperspectral images in the visible and near-infrared region (400–1000 nm) were collected six months after inoculation. The disease incidence was scored with traditional methods based on the presence or absence of visible stem galls. The seedlings were segmented from the background by thresholding normalized difference vegetation index (NDVI) images, and the delineation of individual seedlings was achieved through object detection using the Faster RCNN model. Plant parts were subsequently segmented using the DeepLabv3+ model. The trained DeepLabv3+ model for semantic segmentation achieved a pixel accuracy of 0.76 and a mean Intersection over Union (mIoU) of 0.62. Crown pixels were segmented using geometric features. Support vector machine discrimination models were built for classifying the plants into diseased and non-diseased classes based on spectral data, and balanced accuracy values were calculated for the comparison of model performance. Averaged spectra from the whole plant (balanced accuracy = 61%), the crown (61%), the top half of the stem (77%), and the bottom half of the stem (62%) were used. A classification model built using the spectral data from the top half of the stem was found to be the most accurate, and resulted in an area under the receiver operating characteristic curve (AUC) of 0.83.

Keywords: plant imaging; computer vision; forestry; disease discrimination; hyperspectral imaging; plant phenotyping; machine learning



Citation: Pandey, P.; Payn, K.G.; Lu, Y.; Heine, A.J.; Walker, T.D.; Acosta, J.J.; Young, S. Hyperspectral Imaging Combined with Machine Learning for the Detection of Fusiform Rust Disease Incidence in Loblolly Pine Seedlings. *Remote Sens.* **2021**, *13*, 3595. <https://doi.org/10.3390/rs13183595>

Academic Editor: Thomas Udelhoven

Received: 21 July 2021

Accepted: 3 September 2021

Published: 9 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Pinus taeda L., commonly known as loblolly pine, is the most important forest tree species in the southern United States and is grown for timber, construction lumber, plywood, and pulpwood. Tree breeding programs working on loblolly pine have focused on selecting families based on phenotypic traits such as tree height, volume, stem straightness, and disease resistance [1]. Fusiform rust, caused by the fungus *Cronartium quercuum* f. sp.

fusiforme (Cqf), is a common and damaging disease affecting this species. The fungus typically infects the stem of a young tree, leading to the creation of tumor-like growths known as “rust galls”. This can lead to the death of the tree or to the creation of a “rust bush”, which hinders growth and makes the tree practically unusable for timber production. The planting of disease-resistant seedlings is the most effective strategy to limit the damage caused by this disease [2].

The assessment of fusiform rust disease resistance is performed either through field testing and the evaluation of young trees [3,4] or through artificial inoculation of seedlings in the controlled environment of a rust-screening facility [5]. The latter is beneficial in that fungal spores are applied evenly throughout the test population, and disease incidence can be evaluated within six to nine months post-inoculation. The seedlings are visually evaluated for the presence or absence of rust galls (see Figure 2(left)). Rust disease resistance at the family level is determined by the percentage of galled seedlings observed within each family [6]. However, as described in [7] with respect to various plant species, visual estimation of disease incidence and severity is highly subjective and prone to human errors, in addition to being a labor-intensive process. Moreover, visual assessment can only be conducted after a period of disease infection when symptoms have sufficiently developed.

An objective and rapid method using imaging technology for disease identification would be beneficial for facilitating the disease screening of loblolly pine seedlings. Hyperspectral imaging, which acquires spatial and spectral information simultaneously, has been successfully employed for the detection of plant disease and stress in multiple plant species [8–11]. In these studies, early onset of disease and stress was detected in multiple crops with varying degrees of success. Because of the presence of both spectral and spatial information, hyperspectral imaging provides the opportunity to analyze spectroscopic information at different spatial scales, varying from plant canopy to organs or tissues. This would be beneficial for the assessment of fusiform rust resistance since the disease results in localized symptoms, specifically affecting the stem and branches of young trees.

Based on the location of apparent symptoms, the spectral data from seedling stems can be expected to contain information enabling the successful classification of plants into diseased and non-diseased groups. On the other hand, disease pathogens can often have complex effects on plant physiology and affect the optical properties of multiple plant parts [12]. This is the reasoning behind the use of spectral data from the entire plant as well as from multiple plant parts, not limited to the part with visual symptoms.

A more controlled approach in data acquisition for disease detection is to collect data only from the plant part that displays disease symptoms. An example of this approach can be found in [13], where the authors used hyperspectral imaging to detect *Sclerotinia sclerotiorum* on four-month-old *Brassica* plants by analyzing data from stems stripped of leaves and branches. Data acquired from a targeted method such as this can be expected to increase the accuracy of disease detection models. The sample preparation process, however, decreases imaging throughput and makes repeated image acquisition from the same plants impossible.

The use of top-view hyperspectral images for the detection of fungal infection in *Pinus strobiformis* was presented in [11]. Classification of infection level was based on vigor ratings, and classification accuracy was found to be higher for the samples at the extreme ends of the rating scale. Recently, Lu et al. [10] used hyperspectral imaging of whole plants to assess the freeze tolerance of loblolly pine seedlings. Partitioning of the seedling image into three longitudinal sections for model development led to improved accuracies compared to the model developed using images of the full-length seedlings. In this study, hyperspectral images of the whole seedlings were collected in situ, and image processing methods were developed to extract spectral data from specific plant parts. To further increase the throughput of the system, multiple seedlings were imaged at a time such that a single hyperspectral image cube contained spectral information from five to seven seedlings.

The selection of regions of interest (ROIs) is a vital step in hyperspectral image analysis since acquired images often contain irrelevant pixels that need to be removed. A common approach is to select one or more image channels that provide suitable contrast for ROI segmentation. This selection can be based on domain knowledge or simple observation. Thresholding or other image processing methods can subsequently be used to obtain segmentation masks. In the case of plant images, vegetation indices such as the normalized difference vegetation index (NDVI) derived from two images at the red and near-infrared wavelength bands are commonly used for the segmentation of vegetation pixels [14,15]. While the thresholding of such spectral ratios can be useful for the segmentation of vegetation pixels from the background, the subsequent discrimination of the segmented plant pixels into different classes based on plant parts is a more challenging task. Methods such as the use of an empirically determined wavelength channel [16], machine learning models based on pixel-wise spectral data [17], and deep learning models [18] have been investigated for this purpose.

In this study, we first applied a threshold to NDVI images for the segmentation of vegetation pixels. Subsequently, deep learning models based on Convolutional Neural Networks (CNNs) were trained and utilized for the semantic segmentation of different plant parts. We extracted Red–Green–Blue (RGB) images from the hyperspectral image cube and trained a semantic segmentation model to discriminate the stem and foliage of the loblolly pine seedlings. In addition, a deep learning-based object detection model was used for the delineation of individual plants from images containing a row of seedlings. After the extraction of spectral data from plant segments, support vector machine (SVM) classification models were trained for the classification of diseased and non-diseased plants.

This paper presents an innovative method of using hyperspectral imaging technology for the screening of loblolly pine seedlings for fusiform rust disease incidence by using deep learning methods for plant delineation and plant part segmentation. The methods and results presented here are based on the hyperspectral data collected six months after fungal spore inoculation, which corresponds to the typical time for visual assessment of symptoms [19]. The specific objectives of the current study are as follows:

1. Development of a hyperspectral image processing pipeline for the extraction of spectral data from specific ROIs in images of loblolly pine seedlings;
2. Evaluation of SVM classification models for the discrimination of diseased and non-diseased seedlings based on the spectral data from specific ROIs from (1).

Preliminary results leading to this article were previously presented at the 2020 Annual International Meeting of the American Society of Biological and Agricultural Engineers [20].

2. Materials and Methods

2.1. Plant Materials

The loblolly pine seedlings evaluated in this study were provided by the North Carolina State University Cooperative Tree Improvement Program. The seedling population included a total of 87 seedlots, comprising 84 half-sibling families pollinated with a common pollen mix composed of pollen from 20 unrelated selections, one checklot that consisted of bulked seed from 10 females that were mated with the pollen mix previously described, and two checklots each comprising a single open pollinated family.

Seedlings were sown in May of 2019 at the North Carolina State University, Horticulture Field Laboratory in Raleigh, North Carolina, into Ray Leach Super Cell Cone-tainers (Stuewe and Sons, Inc., Tangent, OR, USA). Two months after sowing, seedlings were organized into replicates. There were 60 replicates, of which 30 replicates were used in this analysis, each with 84 seedlings comprising one seedling per seedlot (incomplete block design). A resolvable row–column incomplete block design was used to maximize the joint occurrence of family in every row and column position across replicates using the CycDesign software [21]. Seedlings were subsequently spaced out such that every other cell was occupied (half-tray capacity, equivalent to 49 cells), to increase growing space per

seedling and root collar diameter. After spacing, each replicate consisted of two trays, and the row within tray and column within replicate corresponded to the incomplete blocks.

In August 2019, the seedlings were artificially inoculated at the USDA Forest Service Resistance Screening Center (RSC) in Asheville, North Carolina. A broad-based inoculum was used, representing the coastal deployment range of loblolly pine. Inocula were sprayed at a density of 50,000 spores per milliliter using a controlled basidiospores inoculation system, placed in a humidity chamber set for optimal growth with temperatures ideal for fungal infection, then moved to the RSC greenhouses. Three weeks later, the seedlings were transported back to the Horticulture Field Laboratory in Raleigh, NC, USA.

2.2. Hyperspectral Image Acquisition

Hyperspectral scans were collected one week prior to inoculation (mid-August, 2019) and at approximate monthly intervals after inoculation (September 2019 to February 2020). The results presented in this article are based on the hyperspectral images collected in February 2020 from 2521 seedlings included in this experiment. The timing of the data corresponds to approximately six months after inoculation, which is the typical timing for visual assessment in conventional phenotyping.

A line scanning hyperspectral imager (Pika XC2, Resonon Inc., Bozeman, MT, USA) was used for the collection of hyperspectral data in the range of 400 to 1000 nm with a spectral resolution of 1.3 nm. The hyperspectral image cubes had the dimensions $1600 \times n \times 462$, where n is the number of line scans used in creating one data cube and 1600 is the number of pixels for each line. The speed of the conveyor stage and the frame rate of the imager were calibrated to ensure an appropriate aspect ratio in the images created. The imager was kept stationary, and a tray loaded with a single row of seedlings was moved across the field of view of the imager using a custom-made conveyor system operated with an electric motor.

The conveyor system was controlled using an Arduino microcontroller that was programmed for forward and reverse motion. The imager was triggered using the bundled software "Spectronon" provided by the manufacturer. A seedling tray was loaded with five to seven seedlings at a time for a side-view scan before being placed on the conveyor board. Halogen lamps were symmetrically placed for additional lighting and a gray cloth was used as a backdrop to facilitate segmentation of plant pixels. A calibration target (SRT-20-020, Labsphere, Inc., North Sutton, NH, USA) with nominal reflectance of 20% was scanned along with the pine seedlings for the standardization of spectral data. A render of the imaging setup is shown in Figure 1.

2.3. Visual Assessment

Post-inoculation visual scores were used as ground truth data for the classification models and were collected at 6, 26, and 48 weeks post-inoculation. The analysis in this paper is based on the visual assessment conducted at 26 weeks post-inoculation. Based on the presence or absence of rust galls or lesions, seedlings were scored as "diseased", "questionable", and "non-diseased". The seedlings ranked as diseased and non-diseased were the only ones used for the analysis presented in this paper.

Less than 1% of the seedlings (51) were observed to show the early development of rust galls that were not the result of the artificial inoculation but were likely caused by ambient fungal spores given the timing of gall development. The ambient infection was expected since the seedlings were not treated with any fungicides. These galls appeared at the base of the seedlings, whereas the galls caused by the artificial inoculation were observed at the approximate height of the plant-tops at the time of inoculation such that these were always situated on the top half of the stem. This is an expected observation since the succulent tissue amenable to infection by basidiospores is present near the apical meristem at the point of inoculation. The plants with galls at the base of the stem can thus be assumed to have been infected at a date earlier than the date of artificial inoculation,

and these plants were also excluded from the discrimination analysis presented in this paper. A histogram of the distribution of seedling scores is shown in Figure 2.



Figure 1. Imaging setup for hyperspectral image acquisition of loblolly pine seedlings. A line scanning hyperspectral imager was used with a motorized conveyor. Supplemental lighting with halogen lamps and a plain backdrop are shown.

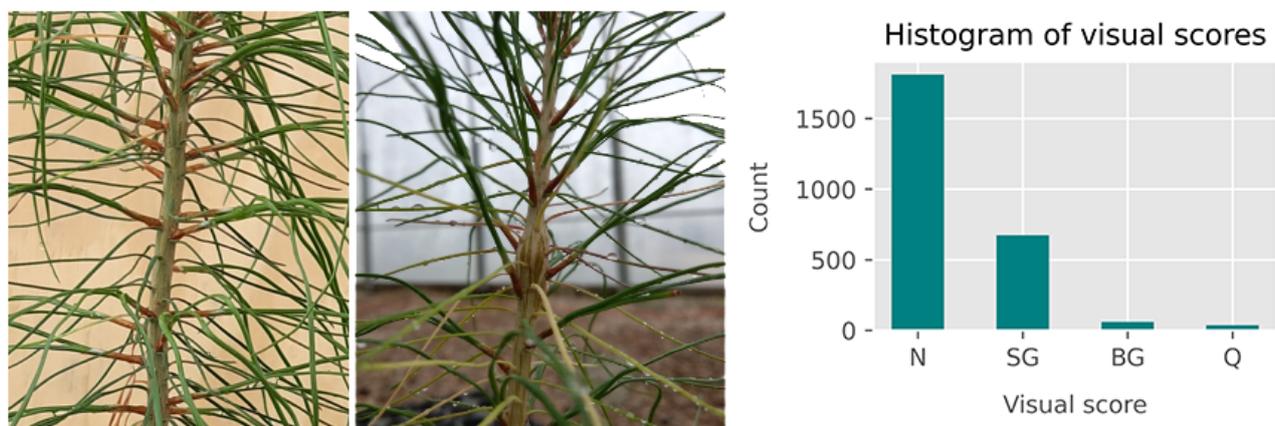


Figure 2. **Left:** Stem of an infected seedling at 3 months (22 November 2019) and 6 months (24 February 2020) after inoculation. The rust gall is clearly visible in the image taken at 6 months to the right. **Right:** Distribution of the visual scores based on the presence or absence of rust galls on the stem. The results are based on scoring conducted at six months after inoculation. N: non-diseased; SG: rust gall on the stem; BG: gall at the base of the plant; Q: questionable.

2.4. Image Processing

2.4.1. Background Removal

Segmentation of plant pixels was carried out by thresholding a normalized difference vegetation index (NDVI) image derived using the two image channels at 705 nm and 750 nm corresponding to the red edge and the near-infrared regions, respectively. This ratio is also referred to as ND705 in the literature; the mathematical relationship used for the calculation is shown in Equation (1) [22,23].

$$NDVI = \frac{I_{750} - I_{705}}{I_{750} + I_{705}} \quad (1)$$

where,

I_{750} = Intensity of the pixel at 750 nm;

I_{705} = Intensity of the pixel at 705 nm.

The plant pixels in the NDVI image were segmented from the background by applying an empirically derived, fixed threshold of 0.1. The initial segmentation was refined by removing noise pixels through area opening, where any connected component with fewer than 1000 pixels was filtered out.

2.4.2. Plant Delineation

The separation of image regions corresponding to individual pine seedlings was necessary for the extraction of spectral data associated with each plant. The separation of seedlings was challenging to achieve because of the variation in heights and orientations combined with the slight overlap between the needles from adjacent plants. A reliable method for estimating the number and location of plants in an image was found to be the presence of the Cone-tainers in which the plants were grown. The consistent shape and size of the Cone-tainers made them suitable features for localizing the plant position using computer vision techniques.

For this purpose, RGB images extracted from the hyperspectral image cube were used to train an object detection and localization model. The wavelength bands used for the red, green, and blue bands were, respectively, 640 nm, 550 nm, and 460 nm. For object detection, the Faster RCNN architecture [24] was selected since it is a general-purpose object detection framework that is computationally efficient and has been shown to perform well in a variety of computer vision tasks. This framework consists of a Fully Convolutional Network (FCN) based on the Region Proposal Network (RPN) that predicts probable image regions containing the object of interest. A pretrained deep CNN is used as the feature extractor before the RPN is deployed in the object detection pipeline. The output of the RPN, which is a set of image regions with an "objectness" score for each region, is used for the extraction of feature vectors that are ultimately used for classification by a fully connected layer. The final output of this model consists of the coordinates of bounding boxes and the probability of an object being contained in each bounding box.

In this study, an implementation of Faster RCNN that is available within the TensorFlow Object Detection API [25,26] with inception v2.0 [27] as the base feature extractor was used for training. A model pretrained with the COCO dataset [28] was selected in order to take advantage of transfer learning. This model was initially fine-tuned for the detection of pine seedling containers using 25 images with 155 annotated instances of the Cone-tainers in the images. The training was conducted by resizing the images to 50% of their size with a learning rate of 0.0002, and training was stopped after 40,000 steps. This trained model was then used to detect the Cone-tainers in 100 additional images, the detected bounding boxes were manually corrected when necessary, and these 100 images were once again used as training data for training a new model. In this case, the training was stopped after 50,000 steps.

The bounding boxes predicted by the object detection model recognizing the Cone-tainers provided a location for the base of the seedlings, but the seedlings were not always vertically oriented. To ensure a better segmentation of the region occupied by each seedling,

the orientation angle of the main stem was estimated by fitting a straight line through the segmented plant pixels after the rudimentary removal of the needles. The color vegetation index [16,29] shown in Equation (2) was calculated and a threshold of 1.9 was used for each pixel.

$$CI = \frac{2 * G}{R + B} \quad (2)$$

where CI is the color vegetation index; R , G , and B are the intensities of the red, green, and blue channels, respectively.

2.4.3. Crown Segmentation

The topmost part of the foliage was segmented out on the basis of the horizontal spread of pixels. Specifically, a horizontal projection or the row-wise sum of non-zero pixels was obtained for the delineated trapezoidal region containing each seedling. The top third section of the segmented plant was taken and the boundary of the crown pixels was determined by finding the most prominent peak in the horizontal projection values. The peak location was determined using the `find_peaks` function provided by the Python library SciPy [30]. The image processing pipeline is shown in Figure 3.

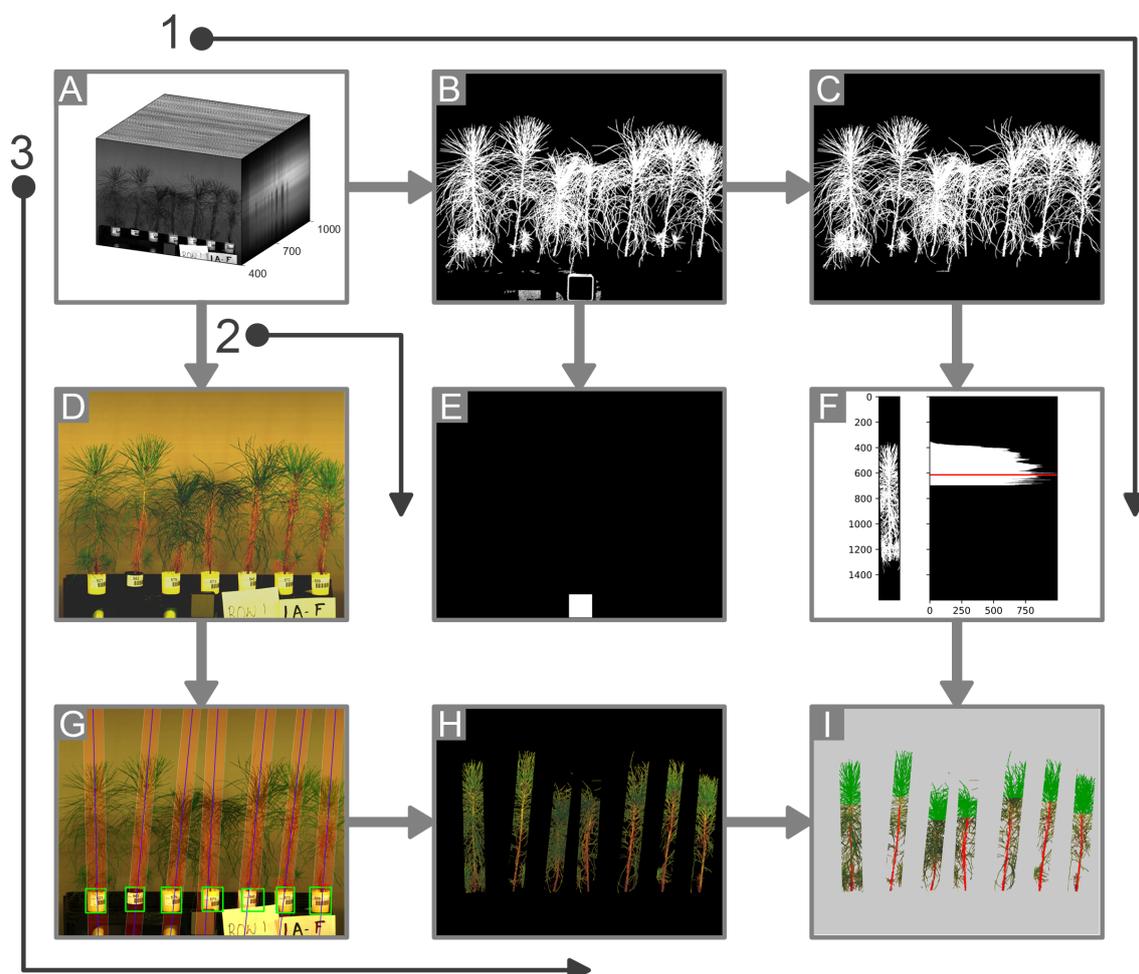


Figure 3. Image processing steps starting with the hyperspectral image cube (A) leading to masks for individual seedlings with segmented crown and stem pixels (I). In process 1, red-edge NDVI image is thresholded to obtain a binary mask (B). Segmented plant pixels are obtained after area opening (C) and a peak-finding algorithm is used to determine the boundary for the crown pixels (F). In process 2, the reflectance standard is segmented out for radiometric calibration (E). In process 3, RGB images (D) extracted from the hyperspectral cubes are used to train a Faster RCNN model used for plant delineation (G,H) and a DeepLabV3+ model used for stem segmentation (I).

2.4.4. Stem Segmentation

Stem segmentation was carried out by using the DeepLabv3+, which is an iteration of the DeepLab CNN model for semantic segmentation [31]. DeepLabv3+ uses an encoder–decoder architecture, where the encoder module consists of the Xception model [32] as the feature extractor. The DeepLab family of segmentation models makes use of atrous or dilated convolution, which uses convolution kernels with gaps between values. This leads to a larger field of view for a kernel without adding to the computational cost that would result from additional parameters in the case of the standard convolution operation. DeepLabv3+ specifically uses Atrous Spatial Pyramid Pooling (ASPP) to encode the features at different spatial scales while using a decoder module for precise delineation of boundaries. The encoder creates a feature map that is a fraction of the size of the input image, and the decoder comes into play for the restoration of the feature map to the original size of the image. The decoder module of the DeepLabv3+ model uses bilinear interpolation for the upsampling combined with convolution operations. The reader is referred to [31] for details on the model architecture; the default model available from the Tensorflow model repository [33] was used with the custom training dataset in this study.

To prepare the training data, a rudimentary stem segmentation was conducted by the thresholding process described in Section 2.4.2. The result of this segmentation was manually corrected using the Image Labeler app included with the Computer Vision Toolbox in Matlab 2019a (Mathworks, Inc., Natick, MA, USA). This approach led to a reduction in the labeling effort since the rudimentary label map was created by the algorithmic approach, and the manual work was limited to the correction of this label map.

The accuracy of the stem segmentation by DeepLabv3+ was assessed using pixel accuracy and mean Intersection over Union (mIoU). Pixel accuracy can be defined as the ratio of the number of correctly classified pixels to the total number of pixels. mIoU is a widely used metric for the evaluation of semantic segmentation models, and it is more robust compared to pixel accuracy in the case of imbalanced class sizes [34]. The mIoU was calculated by taking the average of the Intersection over Union (IoU) values across different classes. The equations used for the calculation of these metrics are shown in Equations (3) and (4).

$$\text{Pixel accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (4)$$

where,

TP = True positive rate;

TN = True negative rate;

FP = False positive rate;

FN = False negative rate.

2.4.5. Segmentation of the Reflectance Standard

An empirical method for the segmentation of the calibration panel was determined based on the spectral properties of the material used to enclose the standard material. The plastic casing was found to be consistently segmented out when using the NDVI thresholding described for the segmentation of the plant pixels in Section 2.4.1. As a result, a group of pixels resembling a square shape was segmented out when thresholding for the plant pixels. This information was used to segment out the pixels of the reflectance standard. Since the same NDVI image could be used for plant segmentation as well as for the segmentation of the calibration panel, this led to a reduction in computation and an increase in the throughput of the segmentation process. The steps involved in the segmentation of the calibration panel can be seen in Figure 3. Other approaches could be used for the segmentation of the reflectance standard—for example, by using the spectral properties of the Spectralon material available from the manufacturer’s calibration. The approach used in this study was adopted solely for convenience and speed of processing.

2.5. Spectral Pre-Processing

The spectra from the pixels in each segmented ROI were averaged to obtain a mean spectrum of raw intensity values. The mean spectrum of the Spectralon reflectance standard (SRT-20-020, Labsphere Inc., North Sutton, NH, USA) was first used for the radiometric correction of raw spectra from the ROIs. The calibration data comprised the reflectance values from 250 nm to 2500 nm for the reflectance standard. The values between 400 nm and 1000 nm relevant for this study were retrieved and linear interpolation was used to obtain reflectance values for each wavelength band in the hyperspectral data of the seedlings. The relationship used for this calculation is shown in Equation (5).

$$R_{ROI} = \frac{I_{ROI}}{I_s} \times R_{FC} \quad (5)$$

where,

R_{ROI} = Reflectance spectrum for the plant region of interest (ROI);

I_{ROI} = Mean raw intensity (i.e., digital number) values for the plant ROI;

I_s = Mean raw intensity values for the Spectralon standard;

R_{FC} = Reflectance values for the Spectralon standard provided by the manufacturer.

Since the extracted dataset included noisy spectra, the Local Outlier Factor [35] with a neighbor size of 10 was used for the removal of outliers. Next, a Savitzky–Golay filter with a window size of 11 and polynomial order 3 was used for spectral smoothing. The window size and polynomial order were selected for efficient smoothing as well as for higher accuracy of the downstream classification models. The dimension of the spectral data was conserved by fitting the polynomial of order 3 to the last window to calculate values for the padded signal. The `savgol_filter` function provided by the Python library SciPy [30] was used for the implementation. Figure 4 shows the processed spectra for the whole plant, with the diseased and non-diseased seedlings differentiated in the figure. The curves represent 1708 non-diseased plants and 653 diseased plants based on the visual scoring.

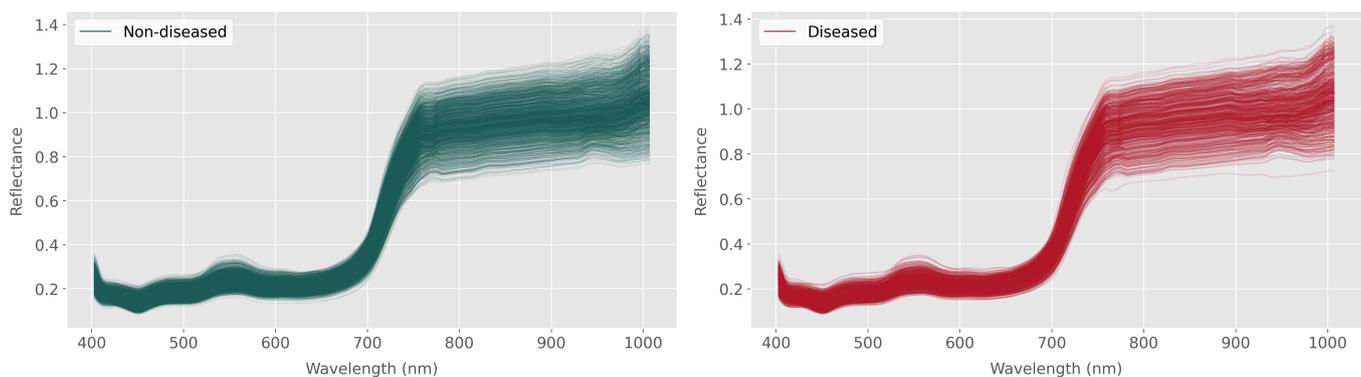


Figure 4. Processed spectra when the whole plant is considered the ROI. The diseased and non-diseased seedlings according to the visual scoring are differentiated. The differences between the diseased and non-diseased groups are not apparent in the plot but machine learning models can learn subtle features, enabling binary discrimination.

2.6. Classification Models

After the processed spectra from each region of interest have been extracted, the discriminant analysis with the extracted spectral data is reduced to the problem of multivariate classification. However, the dataset being imbalanced across the diseased and non-diseased groups (see Figure 2) made it unsuitable for the direct implementation of a general classification algorithm that does not consider class imbalance. Several techniques are available for conducting discrimination analysis with imbalanced datasets, including the use of cost-sensitive learning techniques, and oversampling of the minority class or undersampling of the majority class [36,37]. A widely used method of oversampling cou-

pled with the synthetic generation of data for the minority class is the Synthetic Minority Oversampling Technique (SMOTE) proposed in [38]. SMOTE was implemented for our dataset using the imbalanced-learn library in Python [39].

As part of the investigation into discrimination using the extracted spectral data, support vector machine (SVM) models with a linear kernel were used to classify seedlings into diseased and non-diseased groups. A linear kernel was the preferred choice because of the large number of features, as recommended in [40]. A grid search with 10-fold cross-validation was used to find the optimum value of the regularization parameter C , where the vector used for the search space was [0.1, 1, 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000]. The values for the search were used based on preliminary test runs with the training set. The flowchart in Figure 5 shows the steps followed in the SVM modeling combined with SMOTE oversampling for the classification of diseased and non-diseased groups. The data were first randomly split into a training set (80%) and test set (20%), the training was conducted with 10-fold cross-validation incorporating SMOTE oversampling, and, finally, the selected model was applied to the test set for prediction.

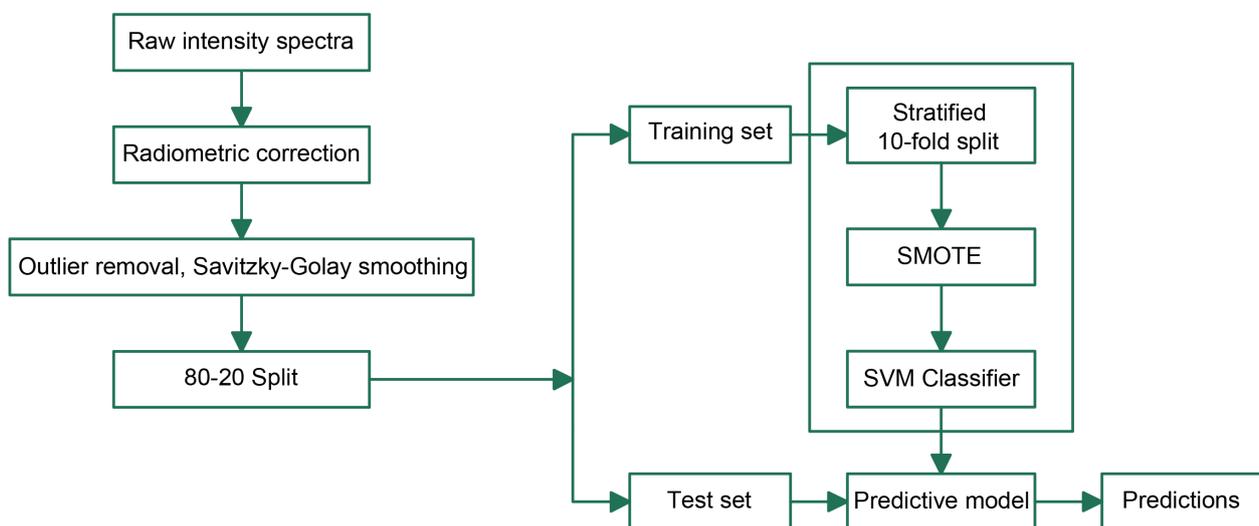


Figure 5. Flowchart showing the steps involved in creating classification model from spectral data. SMOTE stands for Synthetic Minority Oversampling Technique [38].

Balanced accuracy values were calculated as evaluation metrics for the discrimination models. Balanced accuracy is defined as the average of the proportion of correctly classified samples across all classes. In this study, the balanced accuracy of a prediction model provides the average of the accuracy in classification for the diseased and the non-diseased groups. Due to the unbalanced number of samples across the two classes, this is a more robust and reliable measure of accuracy compared to the simple proportion of correctly classified samples. It is also a commonly used metric for the evaluation of plant disease detection models [41,42]. Ten values of balanced accuracy were calculated (with 10-fold cross-validation in each run) for each ROI and a one-way analysis of variance (ANOVA) followed by Tukey's honestly significant difference (HSD) tests at 5% significance level ($p = 0.05$) were conducted to evaluate whether a statistically significant difference existed among the evaluation metrics for the different ROIs.

$$\text{Balanced accuracy} = \frac{TP + TN}{2} \quad (6)$$

where,

TP = True positive rate;

TN = True negative rate.

A commonly used evaluation tool for diagnostic tests is the receiver operating characteristic (ROC) curve [43,44]. The ROC curve is the graph of the true positive rate against the false positive rate. The upper left corner of this plot, representing a 100% true positive rate and a 0% false positive rate, is thus the plot for an ideal classifier. The line inclined at 45° to the positive X-axis, on the other hand, represents a classifier without any power of discrimination. An ROC plot was created, and the area under the curve (AUC) was observed for the optimum ROI selected from the comparison of balanced accuracies. AUC values from the ROC curve are often used as metrics for the evaluation and comparison of binary classifiers [45]. An AUC of 0.5 represents a model equivalent to a random guess, whereas a higher value of AUC represents greater discrimination ability. The use of ROC analysis enables the selection of an appropriate threshold with a desired tradeoff between the true positive rate and true negative rate of a diagnostic test, also referred to as the sensitivity and specificity, respectively. ROC analysis can be especially useful for experiments in plant phenotyping since these experiments are often aimed at finding the best- or the worst-performing samples in the population, and the threshold for discrimination can be adjusted accordingly for the optimum sensitivity or specificity. AUC values have been often used as a metric for the evaluation of plant disease detection algorithms, especially in the case of imbalanced datasets [46,47].

3. Results

3.1. Stem Segmentation

Table 1 shows the pixel accuracy and mIoU values for the segmentation of stem and foliage pixels. mIoU values of 0.69 and 0.62 (with the inclusion and exclusion of background pixels, respectively) indicate a model with acceptable accuracy. Samples of labeled test images paired with the test results are shown in Figure 6.

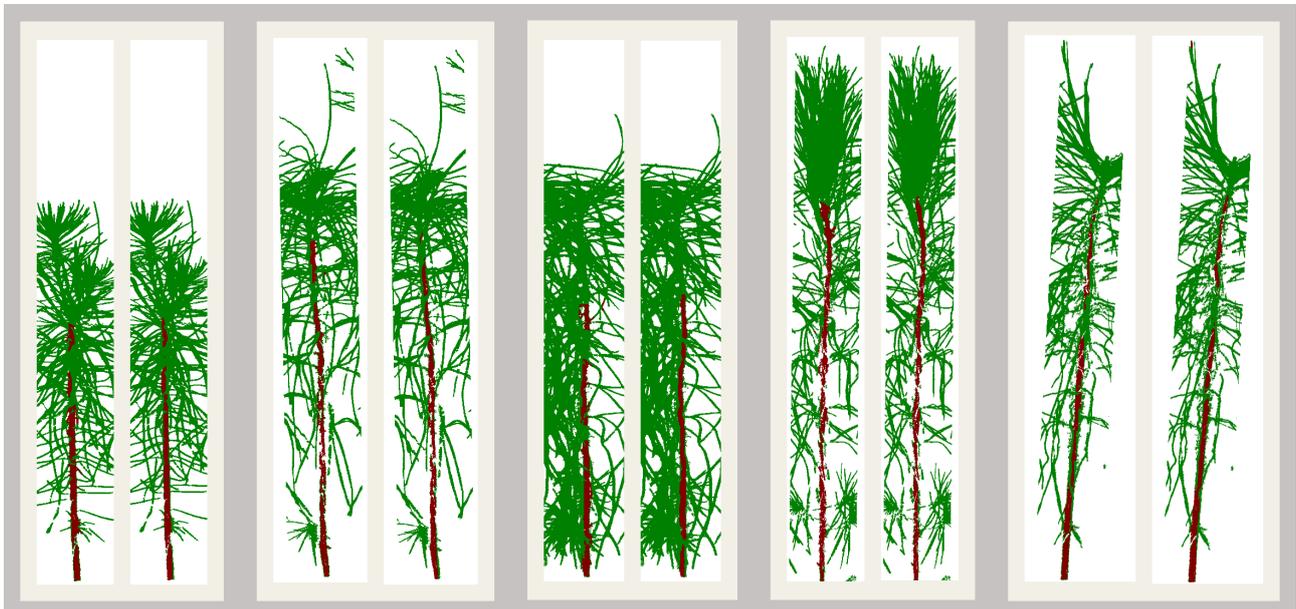


Figure 6. The stem pixels (red) and non-stem pixels (green) for a random group of images from the test set. For each plant, the left image shows the ground truth label and the right image shows the prediction obtained from the DeepLabv3+ model.

Table 1. Pixel accuracy and mean Intersection over Union (mIoU) values for the segmentation of stem and foliage pixels using DeepLabv3+.

	Background Pixels Included	Background Pixels Excluded
Pixel Accuracy	0.83	0.76
mIoU	0.69	0.62

3.2. Classification

The optimum values for the cost parameters found by using the grid search method are shown in Table 2. The corresponding values for balanced accuracy and AUC under the ROC curve are also presented. The balanced accuracy values for the training were derived by averaging the balanced accuracies obtained for each iteration of the 10-fold cross-validation process. The data derived from the stem pixels consistently resulted in higher accuracy of classification of the test set.

Table 2. Results for the classification models for different ROIs. C is the regularization parameter of the SVM models. Accuracies for the training and test sets are balanced accuracies, where the balanced accuracy for training is the average of accuracies obtained from 10-fold cross-validation. AUC is the area under the ROC curve.

ROI	C	Accuracy (Train)	Accuracy (Test)	AUC
Top half of stem	1000	0.76	0.77	0.83
Whole stem	300	0.72	0.71	0.77
Bottom half of stem	400	0.64	0.62	0.66
Crown	1000	0.58	0.61	0.64
Whole plant	500	0.64	0.61	0.67

A pairwise multiple comparison test was carried out to determine if the accuracies across the ROIs were significantly different. Results from ten-fold cross-validation were obtained ten times each for the regions of interest and a Tukey's HSD test was carried out after significant differences were indicated by one-way ANOVA ($F(4,46) = 3137.46$, $p < 0.0001$). Significant differences were observed for the results between every pair.

The boxplots of balanced accuracy values are presented in Figure 7. A distinct and consistent trend is observed when comparing the accuracies across the different ROIs, where the data from the stem pixels result in the most accurate models, and the best model is created using the data from the top half of the stem.

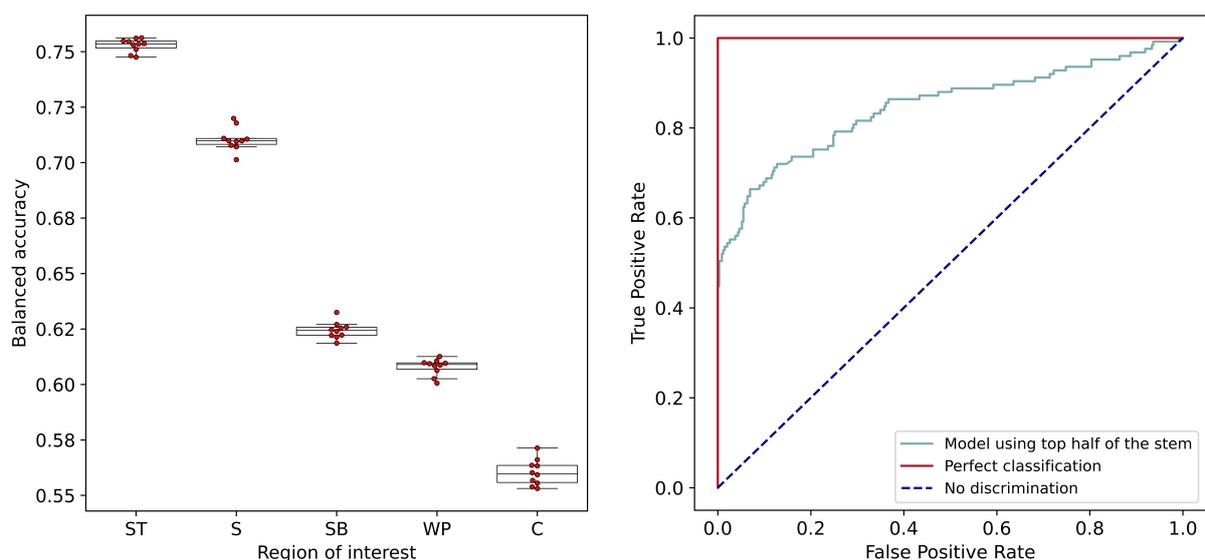


Figure 7. Left: Boxplots showing the balanced accuracies obtained with SVM discrimination models for data extracted from different ROIs. ST: Top half of the stem; S: Whole stem; SB: Bottom half of the stem; WP: Whole plant; C: Crown. Right: Receiver operating characteristic (ROC) curve for the SVM classification model using spectral data from the top half of the stem compared against models with perfect and non-existent discrimination power.

Based on the values of the balanced accuracy, the model using the top half of the stem pixels was selected for further analysis. The ROC curve for the model based on this ROI is also presented in Figure 7.

4. Discussion

In this article, we present a deep learning-based approach for the segmentation of specific regions of interest in Vis–NIR images of loblolly pine seedlings followed by a study in disease discrimination using hyperspectral data extracted from these ROIs. The segmentation of ROIs based on the features extracted from one or more image channels or their combinations is a common technique that has been used in studies using hyperspectral imaging. In addition to the use of traditional techniques of segmentation, such as the red-edge NDVI, the use of the visible section of Vis–NIR images to leverage the power of pretrained deep learning models was found to be a promising approach for the extraction of complex ROIs.

Experiments in plant phenotyping of large populations using hyperspectral imaging have a trade-off between the throughput of data acquisition and the complexity of data processing. The collection of data from well-prepared samples in a controlled environment is ideal for analysis but suffers from low throughput of acquisition. The current study only moderately controlled the data acquisition in a greenhouse environment for the sake of increasing throughput. Images were acquired under ambient lighting with supplemental halogen lamps when necessary, and multiple plants were imaged together in greenhouse trays. As a result, the subsequent processing of the images required multiple steps for the processing and extraction of relevant information. Since a data processing pipeline, once developed, can be repeatedly used for subsequent datasets, we consider a more involved data processing pipeline preferable to the low-throughput acquisition of data.

Morphological symptoms of fusiform rust are visible in the form of galls and lesions on the stem, which makes the detection of symptoms in the images non-trivial. The logical choice of ROI for the acquisition of spectral data was the stem; however, we conducted the study using multiple regions of interest for two reasons:

1. Successful discrimination using the spectral data from the whole plant would increase the efficiency of the data acquisition while also decreasing the effort required in subsequent image processing;
2. Spectral signatures in the NIR region could lead to successful discrimination using spectral data from non-stem pixels irrespective of the location of visible symptoms.

The comparison of discrimination models using data from multiple ROIs indicated that the top half of the stem is the optimum ROI, thus establishing that the region of the visible symptoms should indeed be the ROI of choice. With this conclusion, the image processing pipeline developed for the segmentation of stem pixels will be useful for future experiments and analysis.

When using deep learning models, the labeling of a large number of images can require significant time and effort, especially in the case of pixel-wise labeling for semantic segmentation models. A semi-supervised labeling method was used in this study, where an imperfect but quick segmentation was achieved using a color index to remove most of the non-stem pixels. The “correction” of this imperfect stem segmentation was less labor-intensive compared to the manual labeling of the images.

The discrimination model using the spectral data from the top half of the stem achieved a balanced accuracy rate of 77% and an AUC of 0.83 from the ROC curve. While these values indicate the presence of promising discriminatory power, it is noted that better accuracies are possible and desirable.

One of the conclusions that can be derived from this study is that a more controlled environment for data acquisition, with the targeted imaging of the top half of the stem, would increase the potential for creating more accurate models. A study using the controlled acquisition of hyperspectral images from the infected stem, similar to [13], would be informative for the establishment of a best-case scenario. Since the signals associated

with the disease interact with biotic and abiotic factors, the accuracy and reliability of the detection method can be expected to increase with an increase in the level of control associated with data acquisition.

The use of automated phenotyping platforms with the capability of longitudinal data acquisition in a controlled environment is a promising tool for disease discrimination. As noted by [48], these platforms enable the acquisition of multi-dimensional traits, including changes in physiology and morphology. While the need for visual screening persists for accuracy and reliability [49], the importance of sensor-based discrimination approaches can be expected to increase in the future. In addition to the improvements possible via data acquisition methods, the selection of specific wavelength values with sufficient information for disease detection is another possible avenue of research [42,50]. The reduction of high-dimensional hyperspectral data to spectral indices or spectral signatures derived from a few wavelengths decreases the complexity of data analysis and can also enable the use of less expensive multispectral cameras for data acquisition.

This study represents an initial proof-of-concept for high-throughput fusiform rust disease screening in loblolly pine. Hyperspectral imaging promises several advantages over human-based visual evaluation in terms of its objectivity and amenability to automation, and it can potentially be integrated into a tree improvement selection strategy. In this study, an encouraging level of discriminatory power was achieved using the spectral data from the top half of the stem. The balanced accuracy rate of 77% and an AUC of 0.83 from the ROC curve could be further improved by increasing the level of control in data acquisition and by investigating a wider range of wavelength values. Following further optimization of the scanning environment, there will be the opportunity to evaluate the rate and severity of disease expression caused by the artificial inoculation of seedlings. Family differences could be assessed not solely based on the presence or absence of rust galls but also the rate and severity of gall development. Additionally, hyperspectral data can enable the early detection of disease before the visual symptoms become apparent; this would reduce the time and expense associated with the conventional screening process.

5. Conclusions

We report the use of hyperspectral imaging technology for the detection of fusiform rust disease incidence in loblolly pine seedlings. The study was a novel approach at disease screening, which is normally carried out via visual inspection of symptoms in the greenhouse. We present a workflow incorporating traditional image processing techniques and deep learning methods for the segmentation of plant parts, followed by a support vector machine model for binary classification into diseased and non-diseased samples. We find that this technique is a viable and more efficient technique for the detection of disease incidence. With further work in image acquisition and processing methods, possibly with the use of automated phenotyping platforms, high-throughput phenotyping of loblolly pine seedlings will become an integral part of the methods currently used in resistance-screening centers.

Author Contributions: Conceptualization, K.G.P. and S.Y.; methodology, K.G.P. and S.Y.; software, P.P., J.J.A., and Y.L.; formal analysis, P.P.; investigation, P.P., Y.L., A.J.H., and T.D.W.; resources, K.G.P., A.J.H., and T.D.W.; data curation, P.P., A.J.H., and T.D.W.; writing—original draft preparation, P.P.; writing—review and editing, all; visualization, P.P.; supervision, S.Y. and K.G.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the USDA National Institute of Food and Agriculture, Hatch project 1021499.

Data Availability Statement: Data will be available upon request from the corresponding author.

Acknowledgments: We acknowledge the support provided by the members of the North Carolina State University Tree Improvement Program.

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

References

1. Mullin, T.J.; Andersson, B.; Bastien, J.C.; Beaulieu, J.; Burdon, R.; Dvorak, W.; King, J.; Kondo, T.; Krakowski, J.; Lee, S.; et al. Economic importance, breeding objectives and achievements. In *Genetics, Genomics and Breeding of Conifers*; CRC Press: Boca Raton, FL, USA, 2011; pp. 40–127.
2. Schmidt, R.A. Fusiform rust of southern pines: A major success for forest disease management. *Phytopathology* **2003**, *93*, 1048–1051. [[CrossRef](#)] [[PubMed](#)]
3. McKeand, S.; Li, B.; Amerson, H. Genetic variation in fusiform rust resistance in loblolly pine across a wide geographic range. *Silvae Genet.* **1999**, *48*, 255–259.
4. Spitzer, J.; Isik, F.; Whetten, R.W.; Farjat, A.E.; McKeand, S.E. Correspondence of loblolly pine response for fusiform rust disease from local and wide-ranging tests in the southern United States. *For. Sci.* **2017**, *63*, 496–503.
5. Cowling, E.; Young, C. Narrative history of the resistance screening center: It's origins, leadership and partial list of public benefits and scientific contributions. *Forests* **2013**, *4*, 666–692. [[CrossRef](#)]
6. Kuhlman, E.G. Interaction of virulent single-gall rust isolates of *Cronartium quercuum* f. sp. *fusiforme* and resistant families of loblolly pine. *For. Sci.* **1992**, *38*, 641–651.
7. Bock, C.; Poole, G.; Parker, P.; Gottwald, T. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Crit. Rev. Plant Sci.* **2010**, *29*, 59–107. [[CrossRef](#)]
8. Lowe, A.; Harrison, N.; French, A.P. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* **2017**, *13*, 80. [[CrossRef](#)]
9. Thomas, S.; Kuska, M.T.; Bohnenkamp, D.; Brugger, A.; Alisaac, E.; Wahabzada, M.; Behmann, J.; Mahlein, A.K. Benefits of hyperspectral imaging for plant disease detection and plant protection: A technical perspective. *J. Plant Dis. Prot.* **2018**, *125*, 5–20. [[CrossRef](#)]
10. Lu, Y.; Walker, T.D.; Acosta, J.J.; Young, S.; Pandey, P.; Heine, A.J.; Payn, K.G. Prediction of Freeze Damage and Minimum Winter Temperature of the Seed Source of Loblolly Pine Seedlings Using Hyperspectral Imaging. *For. Sci.* **2021**, *67*, 321–334.
11. Haagsma, M.; Page, G.F.M.; Johnson, J.S.; Still, C.; Waring, K.M.; Sniezko, R.A.; Selker, J.S. Using Hyperspectral Imagery to Detect an Invasive Fungal Pathogen and Symptom Severity in *Pinus strobiformis* Seedlings of Different Genotypes. *Remote Sens.* **2020**, *12*, 4041. [[CrossRef](#)]
12. 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](#)]
13. Kong, W.; Zhang, C.; Huang, W.; Liu, F.; He, Y. Application of hyperspectral imaging to detect *Sclerotinia sclerotiorum* on oilseed rape stems. *Sensors* **2018**, *18*, 123. [[CrossRef](#)]
14. Williams, D.; Britten, A.; McCallum, S.; Jones, H.; Aitkenhead, M.; Karley, A.; Loades, K.; Prashar, A.; Graham, J. A method for automatic segmentation and splitting of hyperspectral images of raspberry plants collected in field conditions. *Plant Methods* **2017**, *13*, 74. [[CrossRef](#)]
15. Asaari, M.S.M.; Mertens, S.; Dhondt, S.; Inzé, D.; Wuyts, N.; Scheunders, P. Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Comput. Electron. Agric.* **2019**, *162*, 749–758. [[CrossRef](#)]
16. Ge, Y.; Bai, G.; Stoerger, V.; Schnable, J.C. Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Comput. Electron. Agric.* **2016**, *127*, 625–632. [[CrossRef](#)]
17. Miao, C.; Xu, Z.; Rodene, E.; Yang, J.; Schnable, J.C. Semantic segmentation of sorghum using hyperspectral data identifies genetic associations. *Plant Phenom.* **2020**, *2020*, 4216373. [[CrossRef](#)] [[PubMed](#)]
18. Feng, X.; Zhan, Y.; Wang, Q.; Yang, X.; Yu, C.; Wang, H.; Tang, Z.; Jiang, D.; Peng, C.; He, Y. Hyperspectral imaging combined with machine learning as a tool to obtain high-throughput plant salt-stress phenotyping. *Plant J.* **2020**, *101*, 1448–1461. [[CrossRef](#)] [[PubMed](#)]
19. Young, C.; Minton, B.; Bronson, J.; Lucas, S. *Resistance Screening Center Procedures Manual: A Step-by-Step Guide Used in the Operational Screening of Southern Pines for Resistance to Fusiform Rust (Revised 2018)*; United States Department of Agriculture Forest Service, Southern Region Forest Health Protection; USDA Forest Service Southern Region: Atlanta, GA, USA, 2018; pp. 1–69.
20. Pandey, P.; Payn, K.G.; Lu, Y.; Heine, A.J.; Walker, T.D.; Young, S. High Throughput Phenotyping for Fusiform Rust Disease Resistance in Loblolly Pine Using Hyperspectral Imaging. In *2020 ASABE Annual International Virtual Meeting*; American Society of Agricultural and Biological Engineers: St. Joseph, MI, USA, 2020; p. 1.
21. Whitaker, D.; Williams, E.; John, J. *CycDesigN: A Package for the Computer Generation of Experimental Designs*; University of Waikato: Hamilton, NZ, USA, 2002.
22. Gitelson, A.; Merzlyak, M.N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol. B Biol.* **1994**, *22*, 247–252. [[CrossRef](#)]
23. Sims, D.A.; Gamon, J.A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* **2002**, *81*, 337–354. [[CrossRef](#)]
24. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)]

25. Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I.; Wojna, Z.; Song, Y.; Guadarrama, S.; et al. Speed/accuracy trade-offs for modern convolutional object detectors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7310–7311.
26. Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I.; Wojna, Z.; Song, Y.; Guadarrama, S.; et al. Tensorflow Object Detection API. 2017. Available online: https://github.com/tensorflow/models/tree/master/research/object_detection (accessed on 15 March 2020).
27. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
28. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 740–755.
29. Meyer, G.E.; Neto, J.C. Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric.* **2008**, *63*, 282–293. [[CrossRef](#)]
30. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]
31. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
32. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
33. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. DeepLab: Deep Labelling for Semantic Image Segmentation. 2018. Available online: <https://github.com/tensorflow/models/tree/master/research/deeplab> (accessed on 20 November 2020).
34. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput.* **2018**, *70*, 41–65. [[CrossRef](#)]
35. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, TX, USA, 15–18 May 2000; pp. 93–104.
36. Japkowicz, N.; Stephen, S. The class imbalance problem: A systematic study. *Intell. Data Anal.* **2002**, *6*, 429–449. [[CrossRef](#)]
37. Krawczyk, B. Learning from imbalanced data: Open challenges and future directions. *Prog. Artif. Intell.* **2016**, *5*, 221–232. [[CrossRef](#)]
38. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
39. Lemaître, G.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* **2017**, *18*, 559–563.
40. Hsu, C.W.; Chang, C.C.; Lin, C.J. *A Practical Guide to Support Vector Classification*; Department of Computer Science, National Taiwan University: Taipei, Taiwan, 2016.
41. Picon, A.; Seitz, M.; Alvarez-Gila, A.; Mohnke, P.; Ortiz-Barredo, A.; Echazarra, J. Crop conditional Convolutional Neural Networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. *Comput. Electron. Agric.* **2019**, *167*, 105093. [[CrossRef](#)]
42. Mahlein, A.K.; Rumpf, T.; Welke, P.; Dehne, H.W.; Plümer, L.; Steiner, U.; Oerke, E.C. Development of spectral indices for detecting and identifying plant diseases. *Remote Sens. Environ.* **2013**, *128*, 21–30. [[CrossRef](#)]
43. Akobeng, A.K. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr.* **2007**, *96*, 644–647. [[CrossRef](#)]
44. Zweig, M.H.; Campbell, G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin. Chem.* **1993**, *39*, 561–577. [[CrossRef](#)]
45. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
46. Johannes, A.; Picon, A.; Alvarez-Gila, A.; Echazarra, J.; Rodriguez-Vaamonde, S.; Navajas, A.D.; Ortiz-Barredo, A. Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Comput. Electron. Agric.* **2017**, *138*, 200–209. [[CrossRef](#)]
47. Clohessy, J.W.; Sanjel, S.; O’Brien, G.K.; Barocco, R.; Kumar, S.; Adkins, S.; Tillman, B.; Wright, D.L.; Small, I.M. Development of a high-throughput plant disease symptom severity assessment tool using machine learning image analysis and integrated geolocation. *Comput. Electron. Agric.* **2021**, *184*, 106089. [[CrossRef](#)]
48. Simko, I.; Jimenez-Berni, J.A.; Sirault, X.R. Phenomic approaches and tools for phytopathologists. *Phytopathology* **2017**, *107*, 6–17. [[CrossRef](#)]
49. Bock, C.H.; Barbedo, J.G.; Del Ponte, E.M.; Bohnenkamp, D.; Mahlein, A.K. From visual estimates to fully automated sensor-based measurements of plant disease severity: Status and challenges for improving accuracy. *Phytopathol. Res.* **2020**, *2*, 9. [[CrossRef](#)]
50. Wei, X.; Johnson, M.A.; Langston, D.B.; Mehl, H.L.; Li, S. Identifying Optimal Wavelengths as Disease Signatures Using Hyperspectral Sensor and Machine Learning. *Remote Sens.* **2021**, *13*, 2833. [[CrossRef](#)]