

Assessing the efficacy of machine learning techniques to characterize soybean defoliation from unmanned aerial vehicles

Zichen Zhang ^{a,*}, Sami Khanal ^b, Amy Raudenbush ^c, Kelley Tilmon ^c, Christopher Stewart ^a

^a Department of Computer Science and Engineering, The Ohio State University, United States

^b Department of Food, Agricultural and Biological Engineering, The Ohio State University, United States

^c Department of Entomology, The Ohio State University, United States



ARTICLE INFO

Keywords:

Soybean leaf defoliation
Convolutional neural networks
Machine learning
Unmanned aerial vehicles
Deep learning

ABSTRACT

Severe crop defoliation caused by insects and pests is linked to low agricultural productivity. If the root cause is not addressed, severe defoliation spreads, damaging whole crop fields. Understanding which areas are afflicted by severe defoliation can help farmers manage crops. Unmanned Aerial Vehicles (UAV) can fly over whole crop fields capturing detailed images. However, it is hard to characterize crop defoliation from aerial images that include multiple, overlapping plants with confounding effects from shadows and lighting. This paper assesses the efficacy of machine learning techniques to characterize defoliation. Given an UAV image as input, these techniques detect if severe defoliation is present. We created a labeled data set on soybean defoliation that comprises over 97,000 UAV images. We compared machine learning techniques ranging from Naive Bayes to neural networks and assessed their efficacy for (1) correctly characterizing images that contain defoliated crops and (2) avoiding wrong characterizations of healthy crops as defoliated. None of the techniques studied achieved high efficacy on both questions. However, we created *DefoNet*, a convolutional neural network designed for detecting crop defoliation that produces models that can be efficacious for either question. If adopted in practice, *DefoNet* models can guide decision making for mitigating crop yield losses due to defoliating insects.

1. Introduction

In the United States, over 500,000 farmers manage soybean crops covering 88 million acres and producing 4.5 billion bushels annually (USDA, 2019). However, a variety of leaf-chewing insects, such as Bean Leaf Beetle, Green Cloverworm, and Japanese Beetle, routinely feed on soybean crops. These insects cause severe *crop defoliation*, i.e., the some discussions about the benefits of network design for *DefoNet* widespread loss of leaf area, which has been linked to loss in agricultural productivity (Haile et al., 1998; Thomas et al., 1974; Higley, 1992; Li et al., 2006).

There exist well-researched treatment threshold guidelines for soybean, which inform the farmer when there is enough insect defoliation damage to warrant an insecticide application. However, current best practice for the assessment of soybean defoliation involves manual collection of upper, middle, and lower-canopy leaf samples from 50 plants per acre (Hunt, 2007). While widely used in practice, this

approach is laborious, tedious and slow. Also, it is very difficult to accurately judge defoliation levels by eyes. Thus, it is very common for farmers and crop consultants to over-estimate both the extent of soybean defoliation and its potential impact on yield, which results in over-treatment and unnecessary expense which decreases farm profit (Man-andhar et al., 2020). Furthermore, the over use of insecticides raises serious problems, from long-term environmental damage to the health of agriculture workers to evolutionary resistance. A second problem is that there are multiple sources of variability within a crop field, such as topography, moisture and soil type, that can affect crop defoliation. So, the collection of a few samples from a few sections of the field can produce unreliable estimates of soybean defoliation. Severity of defoliation in areas where leaves were not collected may be worse than expected. A third problem is that with so many acres to scout, detecting problem levels of defoliation on the ground via manual approaches can be difficult and time consuming. As such, active infestations can spread rapidly while farmers contract, deploy and wait for manual crop

Abbreviations: DL, Deep learning; CNN, Convolutional neural networks; ML, Machine learning; UAV, Unmanned aerial vehicles.

* Corresponding author.

E-mail addresses: zhang.9325@osu.edu (Z. Zhang), khanal.3@osu.edu (S. Khanal), raudenbush.3@osu.edu (A. Raudenbush), tilmon.1@osu.edu (K. Tilmon), cstewart@cse.ohio-state.edu (C. Stewart).

<https://doi.org/10.1016/j.compag.2021.106682>

Received 6 July 2021; Received in revised form 16 November 2021; Accepted 30 December 2021

Available online 20 January 2022

0168-1699/© 2022 Elsevier B.V. All rights reserved.

scouting to finish. Consequently, there is a need for effective, efficient, and nondestructive approaches for detecting and mapping defoliation, and characterizing patterns of disturbance to guide better management decisions about the use of insecticide.

Remote sensing technologies such as UAV provide cost-effective approaches to acquire timely crop health information over a large geographic area that can be used for identifying and mapping soybean defoliation. UAV can fly to preset locations in a field, hover meters above ground, and capture detailed images. Unlike manned airplanes, UAV are piloted via remote control, making them more cost effective for scouting crop fields (Boubin et al., 2019). By covering the whole field, UADeep learning DLConvolutional neural networks CNNMachine learning MLUnmanned aerial vehicles UAV AbbreviAV images can provide holistic views that manual scouting can't. While physiological and phenological crop traits, such as biomass, height, and greenness, can be characterized directly from UAV images (Bendig et al., 2014; Yeom et al., 2019; Larinaga and Brotons, 2019; Anthony et al., 2014), crop defoliation presents challenges. First, aerial images capture multiple plants that overlap in dense canopies, making it hard to mimic manual approaches that analyze individual leaves from specific, isolated crops. Second, aerial images confound defoliation with shadows. To distinguish the effects of lighting from actual crop damage, experts use contextual data based on the field, phenological crop traits, and defoliation patterns in the local region.

Recent research has extensively used machine learning to address complex agricultural problems. For example, Mueller et al. predicted soil erosion using logistic regression (Mueller et al., 2005). The paper concluded that logistic regression has potential for developing erosion indices with soil survey. Lu et al. predicted agricultural water usage using support vector machine (Lu et al., 2009). Grbić et al. deployed multiple Gaussian process regression models to predict stream water temperature with RMSE around 0.87 C and MAE below 0.7 C (Grbić et al., 2013). Fletcher et al. used random forest models fed with leaf multispectral data to classify pigweed and soybean crops (Fletcher and Reddy, 2016). They achieved an overall accuracy of 96.7%. Venkatesh et al. counted corn stalks using CNN models with UAV collected images (Venkatesh et al., 2019). Similarly, studies used deep learning method, such as CNN, for identification of plant disease (Pourreza et al., 2015; Fuentes et al., 2018; Ferentinos, 2018; Barbedo, 2018), as well as for the automated recognition of plants based on morphological patterns of crop leaves (Lee et al., 2015; Grinblat et al., 2016). Although machine learning has been applied to many agricultural related problems, their applications to soybean defoliation assessment have been very limited.

Prior studies such as O'Neal et al. proposed a method based on the analysis of digital images (O'Neal et al., 2002). They collected physical leaf samples and used a desktop scanner to measure leaf area and defoliation. Liang et al. used color-image analysis to estimate soybean leaf area, edge, and defoliation (Liang et al., 2018). The R2 and root mean square error (RMSE) of estimated and observed defoliation of trifoliate leaves were 0.90 and 6.16%, respectively. Da Silva et al. trained CNN models to predict soybean leaf defoliation (da Silva et al., 2019). The data set used to train CNN models was built with single leaf images, and leaf defoliation was manually created by placing irregular shapes on leaves to simulate different degrees of defoliation. The best result achieved a root mean square error of 4.57%. However, the CNN model was trained with only synthetic images of leaves in isolation. While these results are promising, approaches that assess defoliation for individual leaves can not be applied directly to aerial images that include multiple occlusions and overlapping leaves.

The overall objective of this paper is to produce a process that finds which machine learning techniques work well on aerial images collected by UAV for assessment of soybean defoliation. Since there exist several machine learning techniques, the paper compares the efficacy of multiple machine learning algorithms to characterize defoliation. Often times the agricultural data set including soybean defoliation images are imbalanced and thus, it is very challenging to achieve high accuracy in

model performance. Also, the model performance is biased towards classes that are in majority. Although this is a very common issue while applying machine learning techniques in agricultural data set, techniques to address class imbalance to counter bias has received little attention. This paper also assesses techniques to enhance the accuracy of machine learning technique when training and testing sets are imbalanced.

2. Material and methods

Fig. 1 outlines the process for evaluating the efficacy of machine learning for soybean defoliation. In the next section, each stage has been described, focusing on properties that affected the design of our study.

2.1. Data collection

In August and September 2020, we conducted six UAV missions over five soybean fields in Wooster, Ohio, U.S., spanning the critical growth stages when soybean defoliation is usually observed. The fields were planted in April and eventually harvested in early November.

For each field, we defined 10–20 waypoints, i.e., GPS locations for the UAV to fly and capture aerial visible images. A DJI Matrice 200 UAV, mounted with DJI Zenmuse Z5S visible camera, was used to collect 4 megapixel images at an altitude of 10 meters above ground. At each waypoint, a minimum of five images were collected to ensure collection of high quality images. Throughout this paper, we will refer to the image as RGB, reflecting the structure of red, green and blue pixels for each image.

2.2. Expert labeling

We wrote a script file in Python to automatically crop UAV images. Each UAV image was cropped into small images, with each image of size 108×108 pixels. Fig. 2 provides example images from our data set. In total, we randomly selected 94 UAV images that represented areas in a soybean field where at least 5% defoliation were observed. These images were divided into small images. In total, we had 97,395 cropped images from six UAV missions, which were used for expert labelling.

We used our expertise to label the crops shown in each image as (1) healthy or (2) defoliated (Fig. 2). To be sure, manually labeling images is tedious work. Although we could have employed unskilled laborers (Ipeirotis et al., 2010; Hara et al., 2019), to ensure correct labeling, we recruited two experts, 1) a field entomologist with a minimum of 5 years of experience in soybean defoliation and monitoring, and 2) a third year PhD student with multiple publications in precision agriculture forums and experience in collecting and processing agricultural remote sensing data. The experts also serve as co-authors of this manuscript.

To label as many images as accurately as possible, the field entomologist randomly checked the labels assigned by the least experienced labeller. Images where more than 10% of the visible leaf area was defoliated were identified and labelled as defoliated and all other images were labelled as non defoliated or healthy.

Imbalanced Multiclass Defoliation Dataset: As shown in Table 1, 67,479 images were labelled healthy and 29,916 showed greater than 10% defoliation, a ratio greater than 2-to-1. Similar to other agricultural data set focused on diseases, it is common to expect imbalance between healthy and defoliated soybean crops as they are one of the important cash crops to farmers and thus highly managed to avoid potential defoliation. However, imbalanced data challenges machine learning. First, naive algorithms biased toward healthy labels will predict the common case well, even though the models do not capture the traits of defoliation. The most widely used metric to assess efficacy, i.e., accuracy, is biased if test sets are not skewed toward defoliated crops.

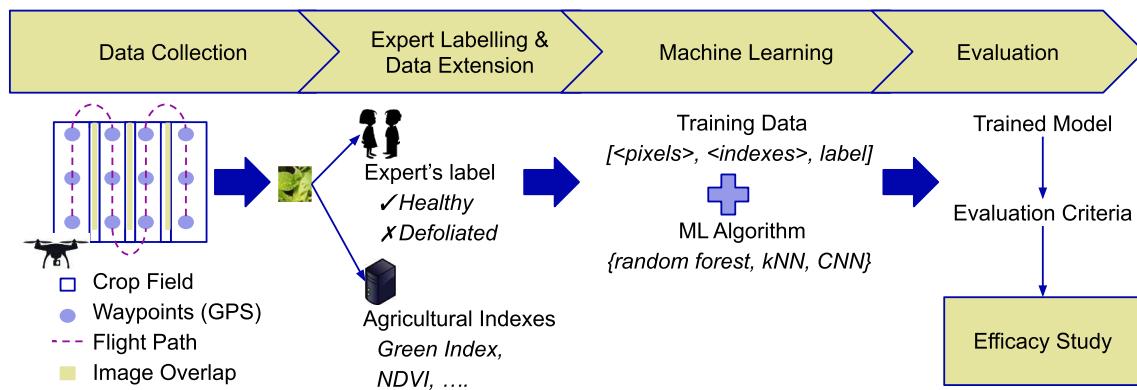


Fig. 1. 4-stage process for a study on the efficacy of machine learning for crop defoliation.

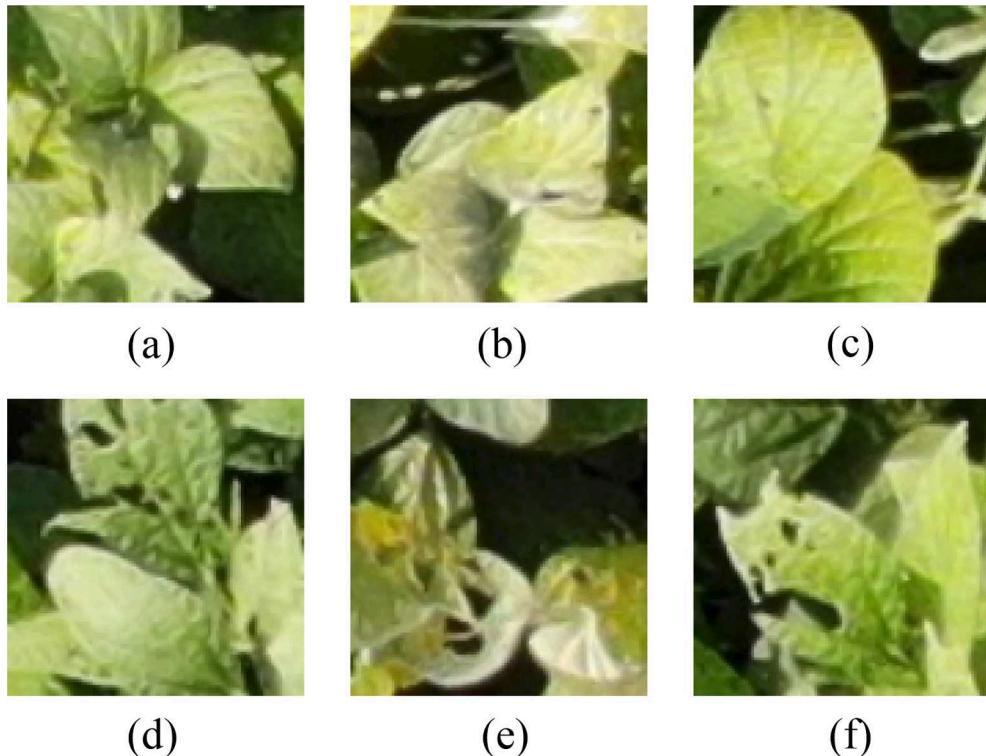


Fig. 2. Example images collected for the study. (a), (b), and (c) were labeled healthy, whereas (d), (e), (f) were labeled as *defoliated*.

Table 1
Healthy and defoliated labels by UAV mission.

Flight No.	Expert Labels	
	Healthy	Defoliated
1	10,353	4,758
2	10,900	4,404
3	9,956	3,654
4	10,575	7,896
5	11,969	6,819
6	10,200	5,911
Total	67,479	29,916

2.3. Data extension

Given a training set, machine learning techniques find patterns to predict the outcome feature (i.e., defoliation in this study). However, machine learning may require very large training sets to learn concepts

easily encoded by experts. These concepts can be encoded as data extensions, reducing the size of the training set required for machine learning. For example, in computer systems, data extensions based on queuing theory make online learning and cloud management feasible (Morris et al., 2018).

Vegetation indexes have been used widely as a proxy for crop conditions (Kogan, 1995; Zhang et al., 2020; Yang et al., 2020; Khanal et al., 2018; Kamble et al., 2013; Holzman et al., 2014). The concept behind vegetation indexes uses arithmetic equations involving spectral signatures (RGB image in this study) at a pixel level. Some vegetation indexes are computed from image data acquired via multispectral or hyperspectral sensors covering visible (VIS) and infrared (IR) regions, which are comparatively expensive than visible sensor. For example, normalized difference vegetation index (NDVI) requires near-IR and red bands. For this work, in addition to RGB images, we selected three vegetation indices, excess green index (ExGI), leaf area to leaf edge index (LAE), and leaf area index (LAI), that can be computed using RGB images, mainly because (1) soybean leaf defoliation can be detected visually,

and (2) these reflect on leaf area which indicates defoliation.

Use of vegetation indexes in our workflow (Fig. 1) reduces the required training set size when vegetation indexes are correlated with defoliation, as well as help better quantify the defoliation.

- ExGI focuses on contrasting green vegetation from soil by giving more weight to green spectral band than red and blue. It can outperform some complex indices that require NIR band (Larrinaga and Brotons, 2019). ExGI equation is as follows:

$$ExGI = (2 \times g) - b - r \quad (1)$$

where $g = G/(R + G + B)$, $b = B/(R + G + B)$, $r = R/(R + G + B)$. R , G , B are spectral bands of RGB images.

- LAE is a ratio of leaf area to leaf edge, and is found to be helpful in characterizing leaf defoliation. For a single leaf, defoliation tends to decrease a leaf area and increase a leaf edge. Thus, the more defoliation, the less LAE is. LAE is calculated by:

$$LAE = \frac{LA}{LE} \quad (2)$$

where LA represents the total leaf area and LE represents the total leaf edge. To calculate LAE , we use ExGI to extract green vegetation and count each pixel for the leaf area. For the leaf edge, we adopt the canny edge detection algorithm to extract the leaf edge and count the number of pixels (Canny, 1986).

- LAI calculates the leaf area from the plant canopy. It is designed to measure the leaf area per unit ground area. Thus, its equation is as follows:

$$LAI = \frac{LA}{GA} \quad (3)$$

where LA is the leaf area while GA is the ground area. The measurement of LA is the same as that in LAE . The ground area is obtained by counting the total pixels of the image.

2.4. Machine learning techniques

We first extracted useful features from aerial images, such as ExGI, LAE, LA, and trained some classic machine learning models. We selected several machine learning classification algorithms that have been widely used in agricultural applications as a start. Furthermore, we also used our data sets to train deep learning approaches, such as VGG16 and ResNet50 deep CNN models, and compared their performance.

- **Naive Bayes Classifier (NB)** (Bhargavi and Jyothi, 2009; Miriti, 2016) consists of a series of simple probabilistic classifiers that use the Bayes theorem under the assumption of strong independence between features.
- **K Nearest Neighbors (KNN)** (Hossain et al., 2019; Suresha et al., 2017) is a supervised machine learning algorithm whose main idea is to classify the k number of nearest data points to a given point as the same class. The classification result varies given k size.
- **Random Forest (RF)** (Lebourgeois et al., 2017; Grimm et al., 2008; Tatsumi et al., 2015) is an ensemble classification model consisting of several decision trees. Its output depends on the majority results given by all decision trees in it.
- **Support Vector Machine (SVM)** (Pujari et al., 2016) can be interpreted as a linear classifier with the largest interval in the feature space. The strategy of SVM is to maximize the interval.
- **Gaussian Process (GP)** (You et al., 2017) is a generalization of Gaussian probability distribution that can be used as a classification algorithm.
- **Convolutional Neural Networks** employ deep learning by adding multiple convolutional layers. Other than *DefoNet* that we custom designed (discussed in Section 2.5), we also trained VGG16 and

ResNet50 models using same data sets for comparisons. VGG16 is a very deep CNN model that achieved 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes (Simonyan and Zisserman, 2014). It consists of 13 convolutional layers and 3 fully connected layers, which in total has over 138 million parameters. ResNet50, short for residual networks, has an architecture that allows us to train extremely deep CNN models with 50 layers (He et al., 2016). The structure of ResNet50 consists of 5 stages each with a convolution and an identity block. For each convolution block, it has 3 convolutional layers and for each identity block, it also has 3 convolutional layers. In total, ResNet50 has over 23 million trainable parameters.

2.5. DefoNet design

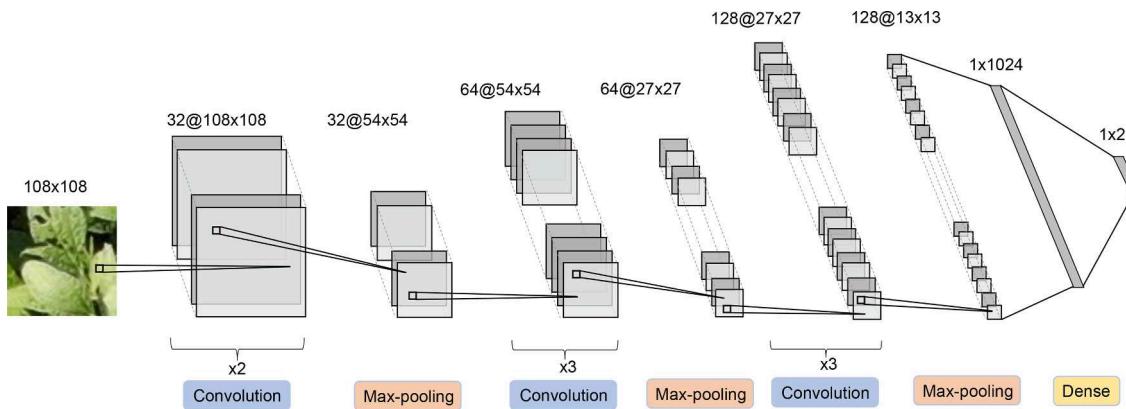
To further explore the potential of deep learning approaches in solving agricultural problems. We designed and fine tuned a CNN model, *DefoNet*, from scratch to test its performance on characterizing soybean leaf defoliation. *DefoNet* is modified from the classical CNN architecture of LeNet (LeCun et al., 1989). The structure of LeNet is very simple, which mainly consists of two convolutional layers. Each convolutional layer is followed by a pooling layer. Based on the structure of LeNet (convolutional layer to pooling layer), we designed a much deeper and more complex CNN model. The idea behind designing a CNN model specific for soybean defoliation is simple. We first started with a simple CNN structure. Based on each result, we added or deleted layers and neurons in each layer accordingly. After the performance of our model stabilized, we tuned model parameters, such as learning rate, number of epochs, regulations, dropout rate, etc. As shown in Fig. 3, the architecture of *DefoNet* has the following features:

- **Input layer.** The size of input images is chosen as 108 * 108 pixels as it is small enough to contain several leaves at a higher resolution and yet not too small to contain a whole leaf.
- **Convolutional layers.** Unlike LeNet that used only two convolutional layers, *DefoNet* is built with a deeper structure that consists of 8 convolutional layers divided into 3 parts. Each convolutional layer uses 3×3 filters, 1 stride, and zero padding. The first part has two convolutional layers with 32 filters for each layer, the second part has three layers with 64 filters for each layer, and the last part has three layers with 128 filters for each layer.
- **Activation layers.** For each convolutional layer, ReLU was adopted as the activation function (Krizhevsky et al., 2012). And we use sigmoid activation function for the last layer of *DefoNet*.
- **Normalization layers.** Each ReLU activation layer is followed by a batch normalization layer to accelerate model training speed (Ioffe and Szegedy, 2015).
- **Pooling layers.** Each convolutional part is followed by a max-pooling layer to down-sample feature maps generated by convolutional layers. The max-pooling has 2×2 pooling size and 1 stride.
- **Dropout layer.** We add a dropout layer with a 0.35 dropout rate before the fully connected layer to avoid model overfitting (Srivastava et al., 2014).

2.6. Techniques to counter imbalance data

In this study, images representing the non defoliated class greatly outnumbered images in the defoliated class (2.25:1 ratio). Thus, a CNN model when trained on a highly imbalanced data naturally tends to perform poorly on the minority class compared to a majority class because there is not enough features in the minority class to learn about. Although the CNN model can yield a decent overall accuracy, it is important to maximize the performance of CNN model for the minority (defoliated) class.

To get a better classification results for the minority class in an imbalanced data set, we adopted two approaches. The first approach is

Fig. 3. The architecture of *DefoNet*.

to increase the class weight of the minority class to balance the total weights between the majority and the minority class so that during the model training process, the classifier will largely weigh the few available minority features. The second approach is to decrease the number of images in the majority class in the training data set by random undersampling. Random undersampling randomly picks a subset of images from the majority class so that two classes will be equal during the model training process. In this way, the classifier can pay the same attention to features in both classes. Random undersampling only happens to the training set of the data set while the test set remains the same. Overall, the original data set, along with two (weight balanced and randomly undersampled) processed data sets were used to train *DefoNet* as well as VGG16 and ResNet50.

2.7. Evaluation method

Models that characterize crop defoliation provide insights for managing crops to avoid defoliation driven yield losses. Efficacy, in this context, means helping farmers learn about soybean defoliation in their field or reduce costs by replacing whole-field insecticide applications with a fewer applications targeting defoliated areas. Thus, models with high accuracy, i.e., over 90% of test images are characterized correctly, are not necessarily efficacious. As areas with severe defoliation are rare, such imbalanced data can allow models to achieve high accuracy by characterizing most images as healthy images. Farmers looking for areas with defoliation may find that such models wrongly characterize healthy images as defoliated, making it hard to understand the true health of the field. Such models thus have limited value for crop management.

Fig. 4 presents a confusion matrix that highlights metrics for efficacy: precision and recall. These metrics thwart imbalance by excluding the dominant label (true healthy). Eqs. 4 and 5 define precision (P) and recall (R).

$$\text{Precision} = \frac{\text{TrueDefoliated}}{\text{TrueDefoliated} + \text{FalseHealthy}} \quad (4)$$

$$\text{Recall} = \frac{\text{TrueDefoliated}}{\text{TrueDefoliated} + \text{FalseDefoliated}} \quad (5)$$

		Predicted Field Conditions	
		Healthy	Defoliated
Actual Field Conditions	Healthy	True Healthy	False Healthy
	Defoliated	False Defoliated	True Defoliated

—→ Recall —→ Precision

Fig. 4. A confusion matrix for leaf defoliation. Positive means defoliated while negative means non defoliated.

Models that achieve high precision help guide farmers to areas with severe defoliation than guiding to areas that are incorrectly classified as defoliated. Models that achieve high recall help find most of the areas with severe defoliation.

We evaluated precision and recall when comparing the performance of the machine learning techniques described in Section 2.4. We used stock implementations from Python Scikit (Pedregosa et al., 2011) to build the models. We split the images from our data set into training and testing subsets. We randomly selected 70% of the images for training and used the remaining images for testing. Experiments were conducted on a server running Windows 10 with Intel(R) Xeon(R) Gold 6258R CPU, 64 GB RAM, and NVIDIA GeForce RTX 2080 Ti GPU.

3. Results

3.1. Machine learning techniques

Fig. 5 reports precision and recall for each of the studied machine learning techniques. Resnet50, VGG16, and *DefoNet* achieve the highest precision and recall. As discussed in Section 2.4, CNNs are well structured for image processing, using deep learning and spatial locality to infer concepts that improve accuracy.

Traditional machine learning techniques achieved lower precision and recall. Naive Bayes achieved 88.4% recall, but only 39.5% precision. Over 65% of healthy images are wrongly characterized as defoliated. In contrast, Random forest achieved 67% precision, but low recall.

Using 2 *DefoNet* models, i.e., one for precision and one for recall, was the most efficacious, achieving 91% precision and 90.9% recall. *DefoNet* reduced false healthy characterizations (precision error) by 1.6X, 1.4X and 3.8X compared to VGG16, ResNet50 and the best traditional machine learning technique respectively. *DefoNet* reduced false defoliated characterizations (recall error) by 1.9X (VGG16), 2.3X (ResNet50) and 1.2X (Naive Bayes) respectively.

Fig. 7 provides a detailed leaf-level comparison of the efficacy of different machine learning models. We selected 12 images that capture

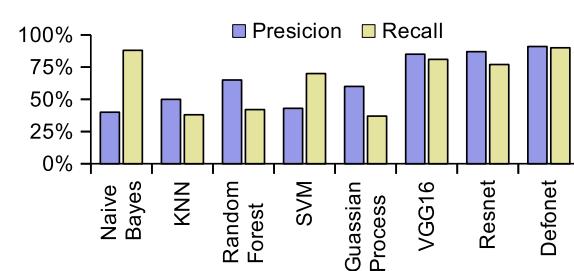


Fig. 5. Precision and recall (efficacy) of competing machine learning techniques.

various challenges with defoliation assessment from UAV. Model predictions on these images were indicative of the limitations of various approaches, especially relative to *DefoNet*. The first two images capture whole, healthy leaves. These images are correctly classified by all but four classifiers. We observed that lighting issues may have upset Gaussian Process and Naive Bayes models on the second image, because these approaches learn distinct non-contiguous classes for images that contain neon green leaves versus images that contain matte green leaves. The third image contains a dark shadow due to depth and lighting. Using only pixels and vegetation indices to split features, the decision tree is unable to extract context that discounts the shadow and over fits. Conversely, random forests use an ensemble of decision trees to remove this bias and correctly predict the third (and fourth) image. The sixth, seventh and eighth images were the hardest for experts to classify. Every machine learning approach struggled to label these images correctly, because they contain overlapping leaves, shadows, actual defoliation, and depth. *DefoNet* (in recall mode) distinguishes these effects, but in precision mode, it reverses its decision on the sixth image, conservatively labeling it as unhealthy. These borderline images separate the efficacy of *DefoNet*. The remaining images capture clear cases of defoliation, but have varying lighting and resolution issues. Deep learning approaches, in general, excel at object detection and, in this case, they identify defoliated leaves well, even under dark lighting that causes simple machine learning approaches to fail (e.g., the eleventh image).

3.2. Impact of imbalanced data

There were changes in recall and precision values among the CNN models trained on the original (imbalanced), class weighted and random undersampling data. Fig. 6 shows results when class weighting and subsampling are used to counter imbalanced data.

Models trained on the original, imbalanced data set, showed higher precision because fewer images were characterized as defoliated. In contrast, models trained using subsamples from training data had the lowest precision. Class weighted models provided good precision and recall. For *DefoNet*, class weighting outperforms training data subsampling in recall while nearly matching the original, imbalanced model on precision.

3.3. Training time

As shown in Table 2, traditional machine learning techniques trained models quickly. Naive Bayes classifier trained a model in 0.03 seconds. Gaussian Process took 1 h and 54 min to train. As shown in Table 3, CNN models trained much more slowly. CNN models trained on the original data set and on the weight balanced data set have similar training time since the two data sets have the same data size. CNN models trained on the randomly subsampled data set have the shortest training time compared to the CNN models trained on the other two data sets because of a smaller data size. From Table 3 we can see that with a smaller CNN architecture, the *DefoNet* models were trained for about 6 times faster

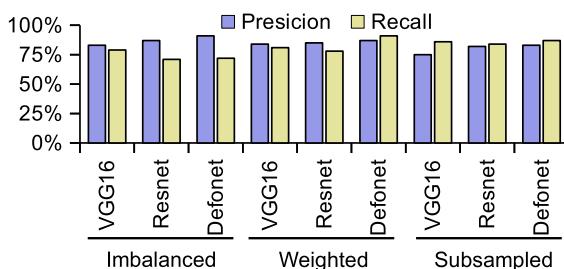


Fig. 6. Comparing VGG16, ResNet50, and *DefoNet* with imbalanced, weighted and subsampled training data.

compared to the ResNet50 models that trained on the same data set, and about 8 times faster compared to the VGG16 models.

3.4. Training data size

We explored the impact of shrinking the training data on (1) training time and (2) efficacy. We repeatedly decreased the training set size by a factor of 2, going from 1/2 to 1/256 (Fig. 8). We randomly selected images for removal. For this experiment, we did not change the size of the testing data set.

We trained *DefoNet* models on each training set with original, imbalanced data. As shown in Fig. 8, precision gets lower as training set gets lower. However recall does not change significantly. With less training data, attempts to find the concepts correlated with defoliated are more error prone, resulting in more False Healthy characterizations and degrading precision. After shrinking the training set below 1/256, recall drops sharply to 46%.

3.5. Image resolution

In addition to changing the data size, we explored the impact of lower resolution imbalanced images on the performance of *DefoNet* models. There are two practical considerations for this experiment. First, in practice, UAV and traditional aircraft often fly at higher altitudes than the altitude used in our study. Second, less expensive UAV may use cameras with a lower resolution.

We used blurring kernels to lower resolution, where larger kernels reflect lower resolution. As shown in Fig. 9, we used kernel sizes of 3×3 , 5×5 , and 7×7 to artificially create images of lower resolution, which were then used to train the *DefoNet* models. We observed that precision has a better tolerance than recall over resolution loss. As we increased the kernel size from 3×3 to 7×7 , recall decreased from 80.6% to 73.8%.

3.6. Multi-class data set

We also explored the potential of the *DefoNet* models on classifying multiple degrees of defoliation severity. For this test, we extended the testing data by labeling images with 20% or greater defoliation. This version of testing data included 3 classes: Class 0 (healthy), Class 1 (10–20% defoliation), Class 2 (greater than 20% defoliation). Again, we observed imbalance innate to the data. 3,818 images were Class 0, 754 images were Class 1, and 57 images were Class 2.

We trained *DefoNet* models on the binary training data to characterize the multi-class testing set. We adapted *DefoNet* by exploiting a feature of CNNs outputs - a probability for each class. We characterized an image as Class 2 if the model outputs a high probability of defoliated and low probability for healthy. We characterized an image Class 1 if the prediction shows that the image was most probable defoliated but healthy probability exceeded a threshold.

Multi-Class Efficacy: Efficacy differs when there are multiple classes compared to only two classes. Here, we will focus on managing very severe defoliation, i.e., Class 2. Areas with very severe crop defoliation call out for manual inspection. Thus, false characterizations as Class 2 can be very costly, causing crop managers to visit healthier areas in the field unnecessarily.

Fig. 10 presents confusion matrices of three *DefoNet* models' results. False characterizations of Class 2 (precision) are rampant for all *DefoNet* models. CNN models with the original and imbalanced training set yield the lowest rate of false Class 2 characterization, which is nearly 7 false labels to 1 one true Class 2. Clearly, *DefoNet* does not achieve efficacy here. However, CNN models trained on weighted and undersampling training sets perform well with regard to true classification of Class 2.

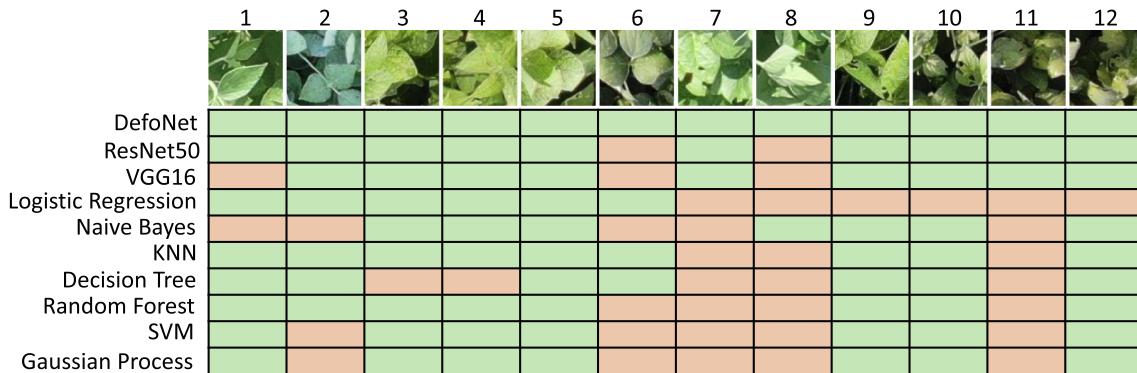


Fig. 7. Prediction result of 12 sample images from machine learning models. First 6 images are healthy while the latter 6 images are defoliated. Green means correct prediction, red means incorrect prediction.

Table 2
The training time of traditional machine learning techniques.

Baseline Model	Naive Bayes	KNN	Random Forest	SVM	Gaussian Process
Time (sec)	0.03	0.42	11.73	263.21	6852.63

4. Discussion

Fig. 11 maps defoliated areas within a soybean field. Each area reflects 108 square pixels, i.e., roughly 7 square inches. If an area in the image depicts severe defoliation according to *DefoNet* (imbalanced), then it has a red bounding box. Characterizations like this can guide decision making for managing crop fields, as evidenced by prior research (Zhang et al., 2020; Khosla et al., 2002). We used the following questions to assess the efficacy of characterizations for decision making.

1. Is decision making affected by *False Healthy*, *False Defoliated*, or both?
2. Does the task demand low error rates (<15%) or does it allow some error (<30%)?
3. Are there time or computational constraints?

4.1. Machine learning techniques

None of the studied machine learning techniques were efficacious for tasks demanding low error rate for *both* precision and recall. However, for many tasks that require either high precision or high recall, CNNs can be efficacious. For example, insurance payouts after unexpected infestations would be based on total damage. Here, characterizations must capture most actual-defoliated areas, so high recall alone suffices. In contrast, insecticide applications today are often applied to whole fields. There are opportunities to reduce costs by applying insecticide to only defoliated areas, requiring high precision to avoid *False Healthy* characterizations.

The design and development of *DefoNet* proved to be an useful way to target the soybean leaf defoliation problem. CNNs can use class weighting to learn efficacious models from imbalanced data. Besides, tuning CNNs for the unique task of detecting defoliated crops can significantly improve accuracy. *DefoNet* was the only technique studied to achieve over 90% accuracy in terms of precision and recall.

Table 3
The training time of CNN models.

CNN Models	DefoNet			VGG16			ResNet50				
	Training Data	imbalanced	weighted	sampled	Training Data	imbalanced	weighted	Training Data	imbalanced	weighted	sampled
Time (min)	505	504	339	4150	4183	2708	3089	3097	2060		

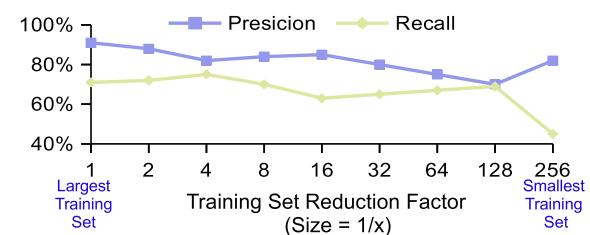


Fig. 8. The performance of *DefoNet* trained on different sizes of training set of the original data set.

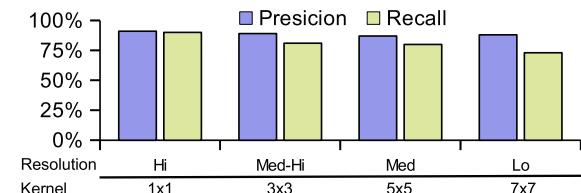
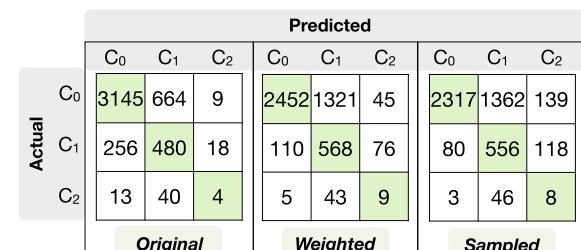


Fig. 9. Performance comparison of models trained on lower resolution data set. Different resolutions are simulated by deploying different sizes of blurring kernels.



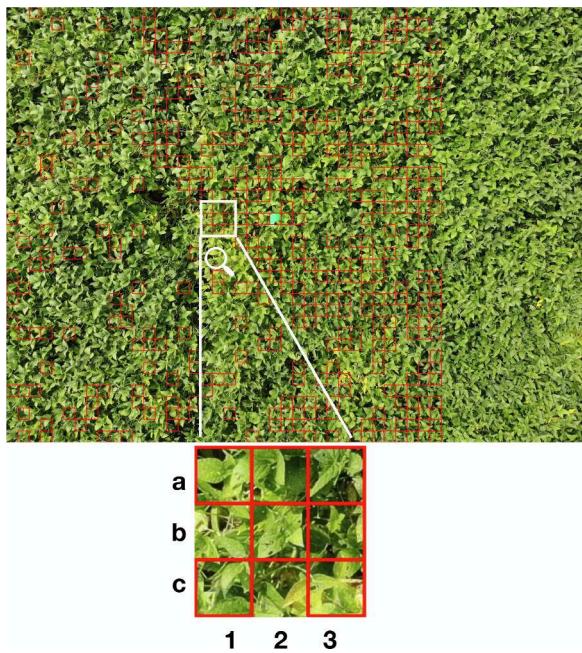


Fig. 11. Aerial image of a soybean field. We used *DefoNet* to map defoliated areas with red bounding boxes. We enlarged a 3×3 block of areas to visualize *DefoNet* characterizations. B1 and C2 were *Healthy* while the other enlarged areas were *Defoliated*.

Furthermore, the process of designing a CNN model, *DefoNet*, for solving a specific agricultural problem can also be used in the future to provide guidance when encountering similar agricultural challenges.

Traditional machine learning techniques produced models quickly. However, in general, these approaches were not efficacious for tasks requiring low recall or precision. These approaches can be useful if models need to be trained quickly for tasks that allow some error. For example, before contracting a professional crop scouting company, a farmer may quickly fly over their field to estimate which services are needed. Here, the farmer can tolerate error as long as the process doesn't take much time. Naive Bayes techniques integrated within the smart-phone app used to set waypoints (e.g., Litchi (Litchi, 2021)) could be efficacious in this example.

4.2. Additional considerations

Training data and Image Resolution: To further explore the robustness of *DefoNet* given different data sizes and image resolutions, we decreased the data size and lowered the image resolution gradually and reported the performance of *DefoNet* models. As shown in Fig. 9, recall dropped slowly while precision didn't lose much performance. This provides options for farmers to balance between cost and outcome when deploying UAV to scout the field. The *DefoNet* model trained on the low resolution data set can lose performance on recognising defoliated images, which lowers recall by classifying more defoliated images as non defoliated. In the meantime, due to the performance loss on classifying the defoliated class, the number of defoliated images that were correctly classified as defoliated and the number of non defoliated images that were misclassified as defoliated both decreased. However, the nominator of the precision equation decreased more compared to that of the denominator. As a result, precision stayed in a relatively stable stage.

Fig. 8 showed the performance comparison of *DefoNet* models trained on different size of data set. In order to make the results comparable, we only decreased the size of the training set. The size of the test set remained unchanged. Decreasing from 100% of the original size to only 8%, recall still performed steadily, wandering around 70%. Despite

the decrease in training size, the performance of *DefoNet* models didn't change much in classifying defoliated images. In the meantime, the ability of *DefoNet* models to learn the features of non defoliated images was compromised; more and more non defoliated images were misclassified, which led to the drop in precision. After decreasing the training size to 1/256, only 182 non defoliated images and 80 defoliated images were left in the training set. At this point, the *DefoNet* model failed to learn enough features of the defoliated class and tended to make predictions towards the non defoliated class, which resulted in the huge performance drop of the *DefoNet* model, especially recall. A recall of merely 46.28% indicate that the *DefoNet* model can only classify around 46% of defoliated images correctly. The increase in precision, however, didn't indicate a performance boost. Instead, the reason for the increase in precision is that both nominator (true positive) and denominator (true positive + false negative) of precision dropped, denominator dropped harder.

Multiclass Defoliation Problem: We explored the potential of deploying *DefoNet* models trained on binary class data set to classify a multiclass defoliation problem. For this, we built a small data set containing three classes: non defoliated, defoliated level 1, and defoliated level 2. Since it's hard to find level 2 defoliated images, the data set is even higher imbalanced, with images 3818, 754 and 57 in non defoliated, defoliated level 1, and defoliated level 2 classes, respectively. Results in Fig. 10 showed a same pattern as the results in Fig. 6 that the *DefoNet* model trained on the original data set obtained the highest precision while the other two *DefoNet* models obtained better recalls. However, due to the introduction of a new class and its insufficient data set, all *DefoNet* models performed poorly on the multiclass data set, especially on classifying defoliated level 2.

5. Limitations

Here, we describe limitations to our study that may be addressed by future work.

Data Collection and Data Expansion: Although all UAV images used to train machine learning models were collected from several dates covering critical growth stages of soybean when defoliation is likely to take place, the fields had very limited areas where defoliation greater than 20% were observed. Future work could extend the data set to include images representing severe soybean defoliations from other fields/regions. Further, UAV images collected using multispectral or thermal cameras could provide better machine learning outcomes. Also, in addition to vegetation indexes, topographical and moisture data can provide important context for assessing defoliation. In particular, we hypothesize that larger data would allow random forests to achieve greater efficacy.

Data Labeling: Our selection of data labellers could yield bias into our process which could lead to results that can not be repeated and do not reflect the true efficacy of machine learning for defoliation. Experts from other regions may use context and crop traits in ways that change image labels. This would affect the marginal efficacy of models. Future work may explore approaches to involve independent experts from other regions or train non-experts in labeling crop defoliation, allowing tools like Amazon Mechanical Turk to produce larger data sets.

Machine Learning Techniques: We assessed and compared several machine learning techniques to characterize soybean leaf defoliation and sought the potential of designing a CNN model as the solution. Other machine learning methods for image recognition, such as semantic segmentation (Girshick et al., 2014; Long et al., 2015), deep random forests, and R-CNN models (He et al., 2015; Girshick, 2015; Ren et al., 2015; He et al., 2017), should also be studied.

Applying Models in Practice As precision agriculture reaches maturity, it is essential to gain experiential studies using model-driven crop management in the field. We are particularly interested in the distribution of practical usage scenarios benefited by precision, recall or accuracy. For example, if most usage scenarios need precision, custom

agricultural UAV and crop scouting models can be developed to maximize throughput and lower cost.

6. Conclusion

In this paper, we seek a solution from machine learning techniques to characterize the severity of soybean leaf defoliation using UAV-collected images. In order to help find the answer to the two key questions proposed in this paper: (1) Are images containing severe defoliation incorrectly characterized as healthy? (2) Do images characterized as healthy contain severe defoliation? The idea is to assess the efficacy of machine learning models from basic machine learning algorithms to state-of-the-art deep CNN models to problem-targeted self-designed CNN model. We design and evaluate the performance of a new CNN model, *DefoNet*, and compare it with many popular machine learning algorithms as well as other CNN models such as VGG16 and ResNet50. Compared to other machine learning models, *DefoNet* can yield a high precision model with 91% accuracy and a high recall model with 90.93% accuracy, which can be adopted for answering either question. The soybean leaf defoliation map generated from UAV images showed that if adopted in practice, *DefoNet* models could aid farmers on getting soybean leaf defoliation condition of their fields in a timely and efficient manner compared to conventional approaches. Furthermore, the process of designing a CNN model, *DefoNet*, for solving a specific agricultural problem can also be used in the future to provide guidance when encountering similar agricultural challenges.

CRediT authorship contribution statement

Zichen Zhang: Conceptualization, Methodology, Software, Writing – review & editing, Data curation. **Sami Khanal:** Conceptualization, Methodology, Software, Writing – review & editing, Data curation. **Amy Raudenbush:** Data curation. **Kelley Tilmon:** Data curation. **Christopher Stewart:** Conceptualization, Methodology, Software, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We received tremendous feedback from Alexander Matveev, Siva R. Meenakshi, Thomas Wenisch and Benjamin Lee. Our work was funded by NSF grants #1749501, #1350941, SNSF NRP75 project Dapprox 407540_167266, Ohio Soybean Council #OSC 20-R-13, and OSU Center for Applied Plant Sciences #013330. We also want to thank Jayson Boubin, Kushal KC, Songyuan Wu for their contributions to the paper.

References

Anthony, D., Elbaum, S., Lorenz, A., Detweiler, C., 2014. On crop height estimation with uavs. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 4805–4812.

Barbedo, J.G.A., 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* 153, 46–53.

Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., Bareth, G., 2014. Estimating biomass of barley using crop surface models (csm's) derived from uav-based rgb imaging. *Remote Sens.* 6 (11), 10395–10412.

Bhargavi, P., Jyothi, S., 2009. Applying naive bayes data mining technique for classification of agricultural land soils. *Int. J. Comput. Sci. Network Sec.* 9 (8), 117–122.

Boubin, J., Chumley, J., Stewart, C., Khanal, S., 2019. Autonomic computing challenges in fully autonomous precision agriculture. In: 2019 IEEE International Conference on Autonomic Computing (ICAC). IEEE, pp. 11–17.

Canny, J., 1986. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.* 6, 679–698.

da Silva, L.A., Bressan, P.O., Gonçalves, D.N., Freitas, D.M., Machado, B.B., Gonçalves, W.N., 2019. Estimating soybean leaf defoliation using convolutional neural networks and synthetic images. *Comput. Electron. Agric.* 156, 360–368.

Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318.

Fletcher, R.S., Reddy, K.N., 2016. Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds. *Comput. Electron. Agric.* 128, 199–206.

Fuentes, A.F., Yoon, S., Lee, J., Park, D.S., 2018. High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. *Front. Plant Sci.* 9, 1162.

Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, p. 580–587.

Grbić, R., Kurtagić, D., Slisković, D., 2013. Stream water temperature prediction based on gaussian process regression. *Expert Syst. Appl.* 40 (18), 7407–7414.

Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on barro colorado islanddigital soil mapping using random forests analysis. *Geoderma* 146 (1–2), 102–113.

Grinblat, G.L., Uzal, L.C., Larese, M.G., Granitto, P.M., 2016. Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* 127, 418–424.

Haile, F.J., Higley, L.G., Specht, J.E., Spomer, S.M., 1998. Soybean leaf morphology and defoliation tolerance. *Agron. J.* 90 (3), 353–362.

Hara, K., Adams, A., Milland, K., Savage, S., Hanrahan, B.V., Bigham, J.P., Callison-Burch, C., 2019. Worker demographics and earnings on amazon mechanical turk: An exploratory analysis. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–6.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Machine Intell.* 37 (9), 1904–1916.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969.

Higley, L.G., 1992. New understandings of soybean defoliation and their implication for pest management. In: Pest Management in Soybean. Springer, pp. 56–65.

Holzman, M.E., Rivas, R., Piccolo, M.C., 2014. Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index. *Int. J. Appl. Earth Obs. Geoinf.* 28, 181–192.

Hossain, E., Hossain, M.F., Rahaman, M.A., 2019. A color and texture based approach for the detection and classification of plant leaf disease using knn classifier. In: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE, pp. 1–6.

Hunt, T., 2007. Evaluating soybean defoliation and treatment need.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456.

Ipeirotis, P.G., Provost, F., Wang, J., 2010. Quality management on amazon mechanical turk. In: Proceedings of the ACM SIGKDD Workshop on Human Computation, pp. 64–67.

Kamble, B., Kilic, A., Hubbard, K., 2013. Estimating crop coefficients using remote sensing-based vegetation index. *Remote Sens.* 5 (4), 1588–1602.

Khanal, S., Fulton, J., Douridas, N., Klopfenstein, A., Shearer, S., 2018. Integrating aerial images for in-season nitrogen management in a corn field. *Comput. Electron. Agric.* 148.

Khosla, R., Fleming, K., Delgado, J., Shaver, T., Westfall, D., 2002. Use of site-specific management zones to improve nitrogen management for precision agriculture. *J. Soil Water Conserv.* 57 (6), 513–518.

Kogan, F.N., 1995. Application of vegetation index and brightness temperature for drought detection. *Adv. Space Res.* 15 (11), 91–100.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Adv. Neural Informat. Process. Syst.* 25, 1097–1105.

Larrinaga, A.R., Brotons, L., 2019. Greenness indices from a low-cost uav imagery as tools for monitoring post-fire forest recovery. *Drones* 3 (1), 6.

Lebougeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., Bégué, A., 2017. A combined random forest and obia classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated sentinel-2 time series, vhr and dem). *Remote Sens.* 9 (3), 259.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1 (4), 541–551.

Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P., 2015. Deep-plant: Plant identification with convolutional neural networks. In: 2015 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 452–456.

Liang, W.-Z., Kirk, K.R., Greene, J.K., 2018. Estimation of soybean leaf area, edge, and defoliation using color image analysis. *Comput. Electron. Agric.* 150, 41–51.

Li, X., An, P., Inanaga, S., Eneji, A.E., Tanabe, K., 2006. Salinity and defoliation effects on soybean growth. *J. Plant Nutrit.* 29 (8), 1499–1508.

Litchi for dji. <https://flylitchi.com/>, 2021.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Lu, S., Cai, Z.-J., Zhang, X.-B., 2009. Forecasting agriculture water consumption based on pso and svm. In: 2009 2nd IEEE International Conference on Computer Science and Information Technology. IEEE, pp. 147–150.

Manandhar, A., Zhu, H., Ozkan, E., Shah, A., 2020. Techno-economic impacts of using a laser-guided variable-rate spraying system to retrofit conventional constant-rate sprayers. *Precision Agric.* 1–16.

Miriti, E., 2016. Classification of selected apple fruit varieties using Naive Bayes. PhD thesis, University of Nairobi.

Morris, N., Stewart, C., Chen, L., Birke, R., et al., 2018. Model-driven computational sprinting. In: ACM Eurosys.

Mueller, T.G., Cetin, H., Fleming, R., Dillon, C., Karathanasis, A., Shearer, S., 2005. Erosion probability maps: Calibrating precision agriculture data with soil surveys using logistic regression. *J. Soil Water Conservat.* 60 (6), 462–468.

ONeal, M.E., Landis, D.A., Isaacs, R., 2002. An inexpensive, accurate method for measuring leaf area and defoliation through digital image analysis. *J. Econ. Entomol.* 95 (6), 1190–1194.

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

Pourreza, A., Lee, W.S., Ehsani, R., Schueller, J.K., Raveh, E., 2015. An optimum method for real-time in-field detection of huanglongbing disease using a vision sensor. *Comput. Electron. Agric.* 110, 221–232.

Pujari, D., Yakkundimath, R., Byadgi, A.S., 2016. Svm and ann based classification of plant diseases using feature reduction technique. *IJIMAI* 3 (7), 6–14.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Machine Learn. Res.* 15 (1), 1929–1958.

Suresha, M., Shreekanth, K., Thirumalesh, B., 2017. Recognition of diseases in paddy leaves using knn classifier. In: 2017 2nd International Conference for Convergence in Technology (I2CT). IEEE, pp. 663–666.

Tatsumi, K., Yamashiki, Y., Torres, M.A.C., Taipe, C.L.R., 2015. Crop classification of upland fields using random forest of time-series landsat 7 etm+ data. *Comput. Electron. Agric.* 115, 171–179.

Thomas, G., Ignoffo, C., Biever, K.D., Smith, D., 1974. Influence of defoliation and depodding on yield of soybeans. *J. Econ. Entomol.* 67 (5), 683–685.

USDA, 2019. United states department of agriculture national agricultural statistic service. 2019. crop production 2018 summary, usda-naas issr 1057–7823.

Venkatesh, S., Stewart, Z.Z.C., Khanal, S., 2019. Corn counting using unmanned aircraft systems and convolutional neural networks. Poster in Department of Food, Agricultural and Biological Engineering.

Yang, M.-D., Boubin, J.G., Tsai, H.P., Tseng, H.-H., Hsu, Y.-C., Stewart, C.C., 2020. Adaptive autonomous uav scouting for rice lodging assessment using edge computing with deep learning edanet. *Comput. Electron. Agric.* 179, 105817.

Yeom, J., Jung, J., Chang, A., Ashapure, A., Maeda, M., Maeda, A., Landivar, J., 2019. Comparison of vegetation indices derived from uav data for differentiation of tillage effects in agriculture. *Remote Sens.* 11 (13), 1548.

You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep gaussian process for crop yield prediction based on remote sensing data. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31.

Zhang, Z., Boubin, J., Stewart, C., Khanal, S., 2020. Whole-field reinforcement learning: A fully autonomous aerial scouting method for precision agriculture. *Sensors* 20 (22), 6585.