ORIGINAL PAPER

Root identification in minirhizotron imagery with multiple instance learning

Guohao Yu[1] · Alina Zare[1] · Hudanyun Sheng[1] · Roser Matamala[2] · Joel Reyes-Cabrera[3] · Felix B. Fritschi[3] · Thomas E. Juenger[4]

Abstract

In this paper, multiple instance learning (MIL) algorithms to automatically perform root detection and segmentation in minirhizotron imagery using only image-level labels are proposed. Root and soil characteristics vary from location to location, and thus, supervised machine learning approaches that are trained with local data provide the best ability to identify and segment roots in minirhizotron imagery. However, labeling roots for training data (or otherwise) is an extremely tedious and time-consuming task. This paper aims to address this problem by labeling data at the image level (rather than the individual root or root pixel level) and train algorithms to perform individual root pixel level segmentation using MIL strategies. Three MIL methods (multiple instance adaptive cosine coherence estimator, multiple instance support vector machine, multiple instance learning with randomized trees) were applied to root detection and compared to non-MIL approaches. The results show that MIL methods improve root segmentation in challenging minirhizotron imagery and reduce the labeling burden. In our results, multiple instance support vector machine outperformed other methods. The multiple instance adaptive cosine coherence estimator algorithm was a close second with an added advantage that it learned an interpretable root signature which identified the traits used to distinguish roots from soil and did not require parameter selection.

Keywords Image segmentation · Multiple instance learning · Minirhizotron image · Image processing · Plant roots · Imprecise labels

1 Introduction

The lack of good sensors, instruments and techniques for field soil measurements is limiting decisive and rapid advances in soil, ecosystem and agronomic science. Currently, only a few techniques are available for visualization of roots in field conditions and those that exist are labor intensive and/or limited in what they can see and measure. Because the root system of plants is involved in the absorption of nutrients and water, understanding their traits, dynamics and behavior is important to a broad number of disciplines, including soil and plant science, agronomists, hydrologist, earth system modelers and others. For example, the study of root systems has become vital as we look for factors that could contribute to the improvement of crop yields to guarantee the food supply. Furthermore, it has been shown that the distribution of roots at depth and their response to fertilizer application substantially affected yield of rice plants [1], and changes in root system architecture, that resulted in improvements in water capture in soil, were directly associated with biomass accumulation and historical yield trends in maize [2]. Similarly, as we look for a sustainable way to manage ecosystem services, there is a critical need to study the root system, as roots are major contributors to the buildup of organic matter in soil, and the role they can play in greenhouse gas mitigation through enhancing the sequestration of carbon in soil.

Common sampling methods used to examine root system dynamics are destructive and provide limited ability to

✉ Guohao Yu
  guohaoyu@ufl.edu

✉ Alina Zare
  azare@ece.ufl.edu

1 Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA

2 Argonne National Laboratory, Lemont, IL 60439, USA

3 Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

4 Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA
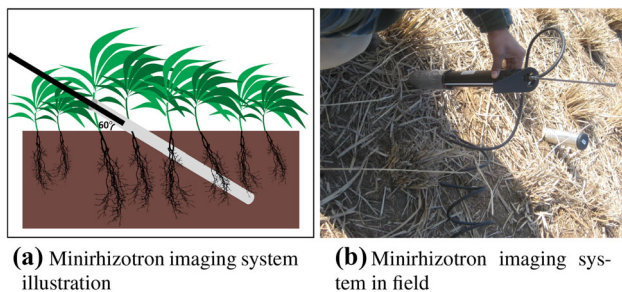
(a) Minirhizotron imaging system illustration

(b) Minirhizotron imaging system in field

Fig. 1 The illustration of a minirhizotron imaging system

draw inferences on the plant response to stresses experienced during the growing season. For example, soil coring, excavation, trenches, or ingrowth cores extract a unique dataset in time that destroys the integrity of the root system and thus limits repeat measurements from the same plant without the introduction of confounding effects [3]. In contrast to these destructive methods, once inserted into the soil, transparent minirhizotron access tubes allow nondestructive assessment of roots over extended periods of time without repeatedly altering critical soil conditions or root processes [4–7]. Roots that grow adjacent to the tube are imaged by inserting a camera (with a light source) into the tube and acquiring images along the length of the tube, providing information about roots present at different soil depths as illustrated in Fig. 1a, b. Since minirhizotron tubes allow for the collection of root images over time, this enables monitoring of complex root growth and turnover dynamics [7]. After collection, minirhizotron images are used to measure and characterize root traits, a process that currently is time-consuming and tedious since standard analysis approaches involve manually outlining and labeling roots using commercially available software such as WinRhizo (Regent Instruments, Canada) or Root-Snap (CID BioScience, Camas, WA, USA).

Machine learning methods have been studied to label roots automatically in minirhizotron imagery. One unsupervised learning approach proposed by Shojaedini and Heidari [8] first initializes a segmentation by partitioning an image using the threshold that maximizes the second-order cross-entropy of the gray-level transition probabilities between the points with intensities above the threshold and those below the threshold [9]. After this initial segmentation, a level set method is used to iteratively improve the identified root boundaries.

In terms of supervised machine learning, Zeng et al. [9] proposed an approach that first extracted a collection of features using linearly shaped spatial filters. Then, the feature map associated with each filter was partitioned into foreground and background classes using local entropy thresholding (LET). Features from each foreground object were then extracted and used within an Adaptive Boosting (AdaBoost) classifier to classify foreground objects as either

root or non-root. Zeng et al. [10] also proposed a rapid root segmentation method that relies on seed point selection. Candidate seed points (identified using a local maxima operation) are classified as root or non-root using a linear classifier. These seed points are then "grown" into roots by identifying a root centerline. More recently, several deep learning methods achieved state-of-the-art results in root segmentation in minirhizotron imagery [11–14]. A number of different neural network architectures have been studied. The U-Net was used as the backbone architecture by Xu et al. and Smith et al. [11,14], Wang et al. [12] relied on the SegNet architecture [42], and hourglass networks [15] were used by [13]. Both Xu et al. and Yasrab et al. investigated transfer learning approaches to overcome small labeled root image dataset [11] and [13]. Yet, these approaches tend to be either sensitive to variation in soil conditions, lighting and root color or require a very large training set. However, labeling individual root pixels to generate a large dataset for a supervised learning algorithm is extremely time consuming, tedious and prone to error.

Although minirhizotrons significantly advance our ability to study plant root systems, data extraction from the acquired images commonly limits the extent of their use as the majority of software available to process images collected in the minirhizotrons requires human differentiation of roots and soil [7]. As described above, efforts to automate the labeling and segmentation of roots in minirhizotrons imagery have been undertaken [8–14,16]. Yet, the best performing of these approaches requires large amounts of training data and it is incredibly labor-intensive, time-consuming, error-prone and tedious to label minirhizotron imagery at the pixel level. In this paper, we propose an efficient approach using MIL methods to label and segment root from minirhizotron data.

MIL algorithms [17–26] only require data to be labeled at the *bag* level. Bags are a multi-set of instances, and each bag is labeled as either "positive" or "negative." A bag is labeled as a positive bag if at least one of the instances in the bag is an instance of the positive target class. If none of the instances in a bag is of the target class, the bag is labeled as a negative bag. The advantage of this framework is that the bags can be constructed in such a way to ease the labeling burden. In our case, the target class corresponds to root. Thus, a positive bag could be an image containing roots and a negative bag is one not containing roots. It is much easier to identify imagery that either does or does not contain roots as opposed to tracing out each individual root segment. MIL algorithms have been used for molecule classification [18–22,27,28], image indexing [25–27]. Among all the MIL algorithms, a few can estimate a discriminative target signature [17,21–26]. The estimated discriminative target signature can be examined to obtain insight into what characterizes the target class and can be easily interpreted to understand what characteristics are

being used to detect and distinguish target from the background.

Our root segmentation approach used training data with only image-level labels instead of full pixel-level annotation. Such an approach significantly reduced the effort associated with label training data for the application of supervised learning algorithms used in the interpretation (or processing) of minirhizotron root images. We propose a downsampling strategy based on histograms of feature computed on superpixels that effectively increased the ratio of target instances to background instances in positive bag. We compare the ability of three MIL methods to detect roots of switchgrass plants over a range of soil textures and colors in minirhizotron images and demonstrated that application of these methods could substantially increase the speed of minirhizotron root segmentation.

2 Methodology

The proposed method consists of three steps: (1) image pre-processing and feature extraction; (2) root detection using an MIL algorithm (multiple instance adaptive cosine coherence estimator, multiple instance support vector machine and multiple instance learning with randomized trees); and (3) post-processing to help reduce false detections.

2.1 Pre-processing and feature extraction

The minirhizotron imagery collected for this study has vertical striping artifacts which is a common occurrence that happens when the numerous sensors in the scan head are slightly differently calibrated resulting in background stripes, or slightly darker or lighter hue bands as shown in Fig. 2a. Pre-processing of the imagery to remove this striping noise [29] consists of subtracting the column mean from each column in the image and adding the global image mean, $\bar{I}$, back to the image as shown in Eq. (1):

$$I_p(m, n, b) = I(m, n, b) - \frac{1}{m}\sum_m I(m, n, b) + \bar{I}(b) \quad (1)$$

where $I(m, n, b)$ is the pixel in the mth row, nth column and bth band of image I, $\bar{I}(b)$ is the global mean of I in band b, and I_p is the resulting pre-processed image. An example of the pre-processed image is shown in Fig. 2b.

After pre-processing, the images are oversegmented into *superpixels* using the simple linear iterative clustering (SLIC) algorithm [30]. Superpixels are groups of pixels which are spatially contiguous and perceptually similar in color as shown in Fig. 2c. This step helps to reduce the overall computational load since processing is done on the reduced
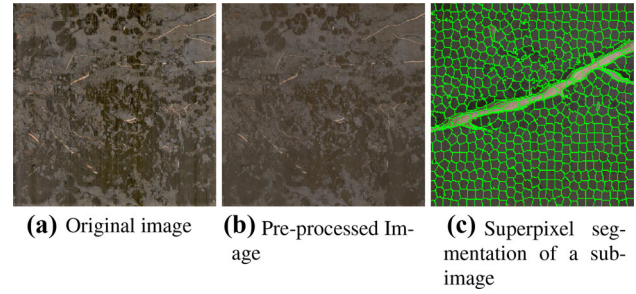

(a) Original image   **(b)** Pre-processed Image   **(c)** Superpixel segmentation of a sub-image

Fig. 2 An example of the image pre-processing sequence: **a** shows the original minirhizotron image, **b** shows the pre-processed image after removing vertical striping noise, and **c** is an example cropped sub-image illustrating the superpixel-segmentation

superpixel level as opposed to the extremely large number of original pixels.

After superpixel segmentation, the mean, variance and entropy of each band in the RGB and LAB color spaces ([mean-R, mean-G, mean-B, mean-L, mean-a, mean-b, var-R, var-G, var-B, var-L, var-a, var-b, H-R, H-G, H-B, H-L, H-a, H-b]) are computed to make an 18-dimensional feature vector for each superpixel. After feature extraction, the superpixel feature vectors of each image are scaled by their order of magnitude so that they are normalized for equal weight across the features,

$$x(i) = \frac{F_i(I_p(m, n))}{s(i)} \quad (2)$$

where F_i is the ith feature generation function, $s(i)$ is the scaling value for the ith feature, and $x(i)$ is the scaled result for the ith feature. These scaled feature vectors are the features used within the MIL algorithms to differentiate between root superpixels and non-root superpixels.

2.2 Multiple instance learning for root segmentation

In our MIL approach, each superpixel is considered an *instance* (or data point) and each minirhizotron image corresponds to a bag. An image containing roots is a positive bag. Otherwise, it is a negative bag. To describe this more precisely, let $\mathbf{B} = \{\mathbf{B}_1, \ldots, \mathbf{B}_K\}$ be the set K bags with label $L = \{L_1, \ldots, L_K\}$. $L_i \in \{0, 1\}$, where 0 represents a negative bag and 1 represents a positive bag. Each bag, $\mathbf{B}_i$ contains N_i superpixels, $\mathbf{B}_i = \left[\mathbf{x}_i^1, \mathbf{x}_i^2, \ldots, \mathbf{x}_i^{N_i}\right]$. Each feature vector, $\mathbf{x}_i^j \in \mathbb{R}^{18}$ is computed on a superpixel as described in Sect. 2.1. A bag, $\mathbf{B}_i$, is labeled as positive, $L_i = 1$, if there exists at least one feature vector in the bag corresponding to a root superpixel.

Three MIL algorithms are investigated for root detection and segmentation, the multiple instance adaptive cosine coherence estimator (MI-ACE) [17], the multiple instance
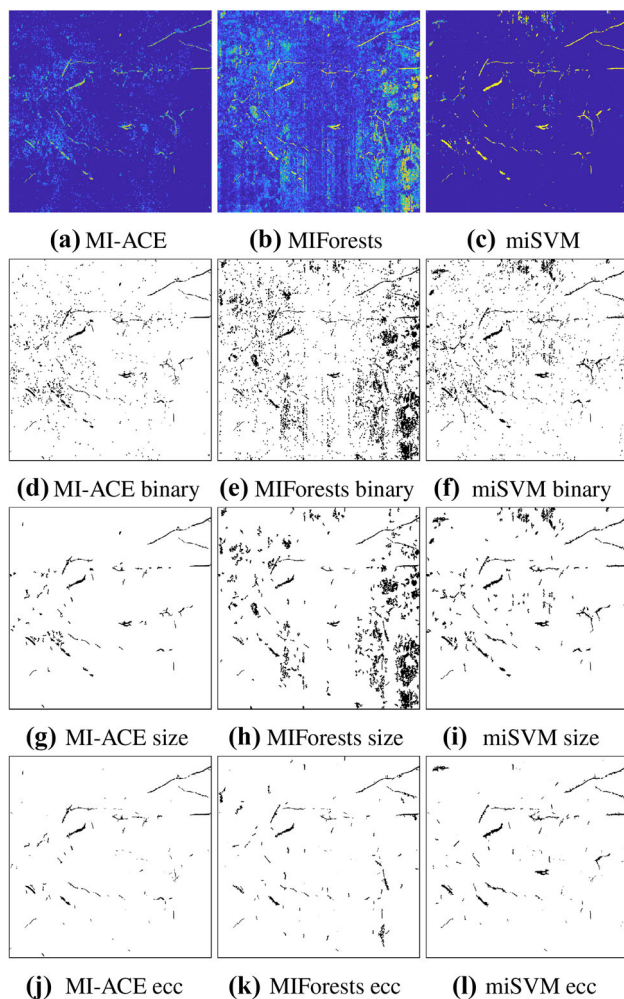
Fig. 3 MIL algorithm results corresponding to the image shown in Fig. 2b. **a–c** show the MI-ACE, MIForests, miSVM confidence maps, respectively. The confidence map is the pixel-level output from the MIL algorithms indicating the confidence the algorithms assign to each pixel for belonging to the root class. The color map for these confidences ranges from dark blue (very low confidence) to bright yellow (very high confidence). **d–f** are binarized results after thresholding MI-ACE, MIForests, and miSVM confidences maps, respectively. **g–i** are the results after filtering the threshold results based on connected component size where components with too few pixels are removed. **j–l** are the results after filtering based on connected component eccentricity where components that do not have sufficiently large enough eccentricity are removed

support vector machine (miSVM) [28], and the multiple instance learning with randomized trees (MIForests) [27] algorithms. MI-ACE is a supervised multiple instance learning algorithm that estimates a discriminative target signature (in this case, root signatures) **s** during training and, then, uses this discriminative signature within the adaptive cosine estimator detector [31–33] to detect root in unlabeled test imagery. The advantage of MI-ACE (and other MIL concept learning methods [21–26]) is that the estimated discriminative target signature can be examined to obtain insight into

what characterizes the target (i.e.,root) class and can be easily interpreted to understand what root characteristics are being used to detect the roots and distinguish them from the background (i.e.,soil). The miSVM algorithm iteratively trains a support vector machine [34] classifier while enforcing the multiple instance learning constraints that each positive bag contains a target and each negative bag does not contain any target. MIForests uses a deterministic annealing approach to optimize an objective function that enforces the MIL constraints on the data while training a random forest classifier. MIForests initially trains a random forest classifier using all data instances, and each instance is labeled the same as its bag label. Then, it iteratively refines the instance labels while optimizing the MIForests objective function.

We trained the three MIL algorithms with pre-processed minirhizotron images with image-level labels. The trained models were then used to predict the roots in other images and to determine the degree of confidence of those predictions as illustrated in Fig. 3a–c (see Sect. 3.2 for details).

2.3 Post-processing

After computing the confidence map, some post-processing is applied to reduce the number of false detections. First, the confidence map is binarized by thresholding as shown in Fig. 3d–f. After thresholding, root shape and size characteristics are used to filter out likely false detections. The size and eccentricity are computed on each connected component made by pixels identified as root. Connected components with a small size and a small eccentricity are removed from the binarized image. The results after post-processing by size are shown in Fig. 3g–i. The results after post-processing by eccentricity are shown in Fig. 3j–l.

3 Experiments

3.1 Data description

Switchgrass (*Panicum virgatum* L.) is recognized as the most promising perennial grass for bioenergy production in the USA. Nonetheless, despite being the subject of abundant and multidisciplinary research, there is a lack of understanding of the contributions of the switchgrass root system to the adaptability to different environments, biomass productivity and substantial ecosystem services [35]. The extensive root system of switchgrass is increasingly recognized to contribute to long-term carbon sequestration, high resource use efficiency and high biomass productivity [36,37]. However, despite the essential biological function, information about switchgrass root system dynamics in space and time is limited, at least in part because of the challenges associated with access to the soil-root interface without disturbing the soil environment,
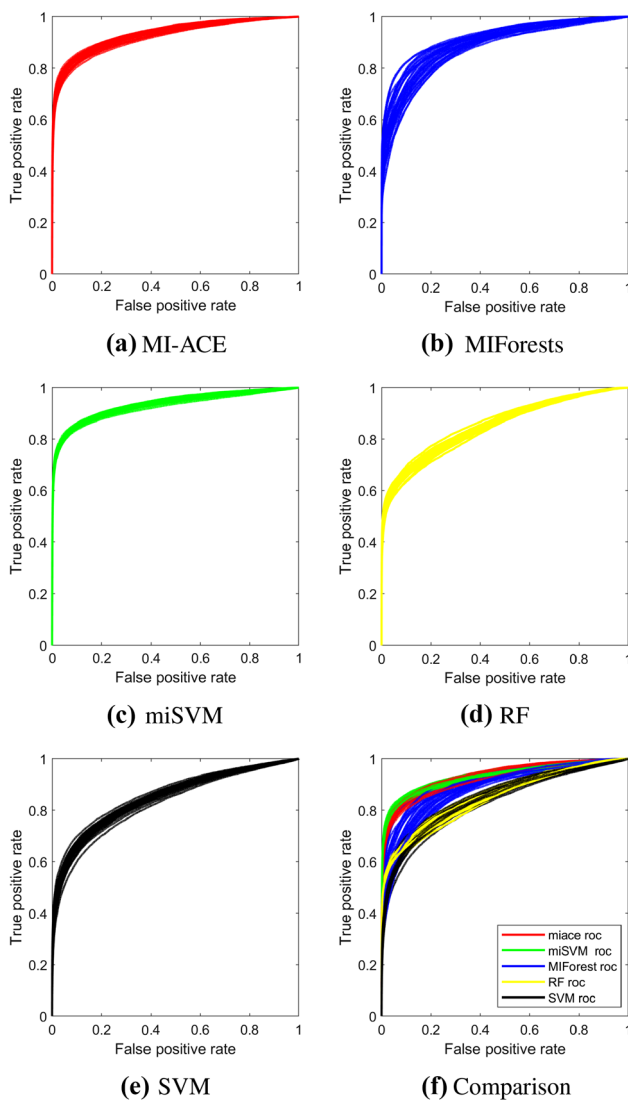
(a) MI-ACE

(b) MIForests

(c) miSVM

(d) RF

(e) SVM

(f) Comparison

Fig. 4 Root detection results with different methods across 30 runs without post-processing. **a** ROC curves of MI-ACE. **b** ROC curves of MIForests. **c** ROC curves of miSVM. **d** ROC curves of RF. **e** ROC curves of SVM. **f** Comparison of ROC curves across different methods

which can impair subsequent measurements of the same plant [38].

Switchgrass root images were collected from minirhizotron access tubes that were installed in a 2-y-old switchgrass field at the US Department of Energy National Environmental Research Park at Fermilab in Batavia, IL, USA. Minirhizotron access tubes measuring 1.82 m in length were installed at 60° off the horizontal axis using an angled, guided hydraulic soil core sampler. The tubes were inserted to reach a maximum vertical depth of approximately 1.2 m, and foam caps were installed on the end protruding from the soil to insulate the tubes, block light and protect them from UV damage. Images were collected by inserting a CI-602 in situ root imager (CID BioScience, Camas, WA, USA) into the minirhizotron tubes. This scanner was set to collect 360-degree images at 300 dpi in 28-cm depth increments.

3.2 Experimental setup

The pixel level segmentation was estimated with MIL algorithms (MI-ACE, miSVM, MIForests) using training imagery with only image level labels. The training data for this study contained 34 positive bags and 34 negative bags. Each bag contains instances from an image, and the bag label is the same as the image label.

To help reduce computational complexity, a subset of the instances in each bag were selected and used within the MIL algorithms. These instances were selected in such a way to ensure a wide distribution of root and non-root instances with varying color properties. Specifically, a 200-bin histogram of the green channel values was constructed using the feature vectors across all superpixels. Then, one superpixel corresponding to each non-empty histogram bin was selected at random (uniform random selection) and kept for processing. If all bins are non-empty, this results in 200 instances per bag. Three multiple instance learning models: MI-ACE, miSVM and MIForests were trained multiple times. Each time, the instances in each bag were randomly selected again using the approach described above and the same training data were used for all three models.

For each method, parameters were tuned to maximize performance. Performance was measured using the F-score [39] on the validation dataset. For the miSVM algorithm, a radial basis function (RBF) kernel was used and the parameters were set to be $C = 10$ and $\gamma = 1$. These parameters were determined by varying the RBF kernel width, γ, from 2^{-15} to 2^3 by a factor of 4. Similarly, the parameter C was varied from 2^{-5} to 2^{15} by a factor of 4. The parameter settings with the best performance over this range were selected as the parameters to be used for comparison. For MIForests, the parameters were set to be $t = 100$ and $x_d = 4$. These parameters were selected by varying the number of trees t from 2^4 to 2^{13} by a factor of 2, and the dimension of the subset of input variables x_d was varied from 2 to 16 by a factor of 2. All other parameters of MIForests are set to be the default value. MI-ACE did not have parameters that required tuning. After training and validation, the models were tested by a set of images with pixel level labels.

3.3 Root detection results

3.3.1 MIL methods

MIL methods (MI-ACE, MIForests and miSVM) were run 30 times, and the root detection results were evaluated by receiver operating characteristic (ROC) curves, shown in Fig. 4a–c. The mean and variance of the true positive rate

(TPR) of different methods at false positive rate (FPR) ranging from 0.01 to 0.06 are presented in Table 1. It was found that miSVM outperformed the other approaches with a higher average TPR at the same FPR, while MI-ACE was a close second. The performance of miSVM was also more consistent than that of the others. Namely, miSVM had the least variance in performance. In contrast, the MIForests variance was one order of magnitude larger than the others, indicating that the MIForests approach was more sensitive to samples selected in each bag than the others.

3.3.2 Comparison with non-MIL methods

The results of MI-ACE, MIForests and miSVM were compared with non-MIL methods (support vector machine and random forests). The training data used to run the support vector machine (SVM) [40] and random forests (RF) [41] methods were the same as those instances in bags used to run miSVM and MIForests, and each instance is assigned the bag label. The parameters of SVM and RF were also set to be the same as miSVM and MIForests. Both SVM and RF were run 30 times. Figure 4d, e shows the ROC curves of RF and SVM which are compared with MI-ACE, MIForests and miSVM in Fig. 4f.

The ROC curves show that non-MIL methods cannot compete with MIL methods when dealing with imprecisely labeled root data (i.e., labeled at the image level). The ROC curves of non-MIL methods were below the ROC curves of the MIL methods which means that at the same FPR value, the average TPR of non-MIL methods was less than the TPR value of MIL methods. For the case of FPR = 0.03 in Table 1, the average TPR of miSVM was 0.78 which detected 26% more roots than the SVM method and 20% more roots than the RF method. The average TPR of MI-ACE was 0.75 which detected 23% more roots than the SVM method and 17% more roots than the RF method. However, the average TPR of MIForests was similar to the average TPR of RF because the variance of MIForests TPR was one order of magnitude greater than that of non-MIL methods, and some of the ROC curves of the MIForests method overlapped with the RF method ROC curves. Therefore, the average performance of MIForests was similar to the RF.

3.3.3 Post-processing results

Results were also evaluated after post-processing. The confidence maps from each image were thresholded at FPR values ranging from 0.01 to 0.05. These binarized maps were constructed by filtering with an eccentricity or size value lower than a fixed threshold to remove connected components. The threshold was varied to produce the ROC curves shown in Fig. 5. In the post-processing step, it was important to decrease the FPR without decreasing TPR too much so
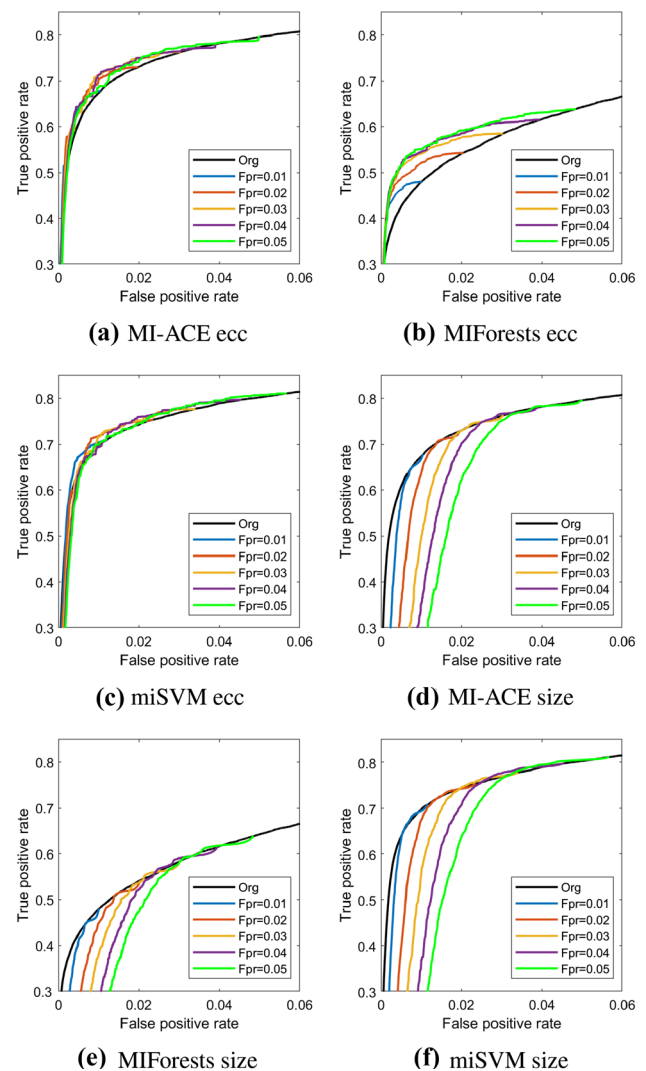

(a) MI-ACE ecc   **(b)** MIForests ecc

(c) miSVM ecc   **(d)** MI-ACE size

(e) MIForests size   **(f)** miSVM size

Fig. 5 ROC curves after post-processing. Results are shown after thresholding confidence maps at FPR rates ranging from 0.01 to 0.05. After thresholding, results are filtered by removing connected components with either too few pixels in size or too small of an eccentricity score. The Org ROC curve is the ROC curve without post-processing. **a–c** show ROC curves of post-process with eccentricity computed from connected components in the binarized confidence maps of MI-ACE, MIForests and miSVM. These ROC curves are generated by varying the eccentricity threshold. **d–f** show ROC curves of post-process with size computed from connected components in the binarized confidence maps of MI-ACE, MIForests and miSVM. These ROC curves are generated by varying the size threshold

that false roots were removed from binarized maps and root objects were not affected. The eccentricity can effectively decrease FPR and keep TPR changing slowly such that at the same FPR, the TPR on post-processing ROC curves using eccentricity was larger than the TPR on Org ROC curves as shown in Fig. 5a–c. This indicated that the root objects had larger eccentricity than non-root objects. Size can also be used to remove false roots. The ROC curves using size to remove likely false roots show that root objects had larger

Table 1 Mean and variance of TPR of comparison algorithms at FPR varying from 0.01 to 0.06

FPR	MI-ACE TPR		MIForests TPR		miSVM TPR		SVM TPR		RF TPR	
	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
0.01	*0.66*	*0.38e-03*	0.49	3.7e-03	**0.70**	**0.21e-03**	0.40	0.87e-03	0.51	0.47e-03
0.02	*0.72*	*0.33e-03*	0.54	4.1e-03	**0.75**	**0.12e-03**	0.46	0.74e-03	0.55	0.33e-03
0.03	*0.75*	*0.32e-03*	0.58	4.1e-03	**0.78**	**0.10e-03**	0.52	0.67e-03	0.58	0.26e-03
0.04	*0.77*	*0.29e-03*	0.61	4.0e-03	**0.80**	**0.08e-03**	0.55	0.61e-03	0.60	0.23e-03
0.05	*0.79*	*0.25e-03*	0.63	3.8e-03	**0.81**	**0.07e-03**	0.57	0.55e-03	0.61	0.20e-03
0.06	*0.80*	*0.23e-03*	0.66	3.6e-03	**0.82**	**0.06e-03**	0.59	0.52e-03	0.63	0.22e-03

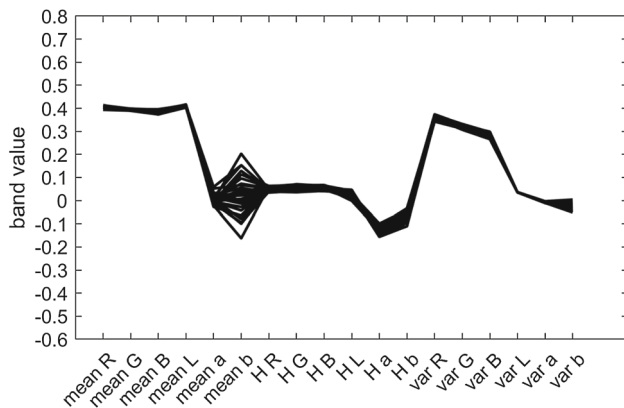Bold reflects the best performance; italics reflects the second best performance

Fig. 6 Root signatures. Root signatures generated by MI-ACE method with training data labeling one image as a bag. The [mean-R, mean-G, mean-B, mean-L, mean-a, mean-b, H-R, H-G, H-B, H-L, H-a, H-b, var-R, var-G, var-B, var-L, var-a, var-b] features correspond to the 18-feature set extracted from each superpixel. They corresponding to the mean, entropy and variance of samples in RGB and Lab band of each image, respectively

size than non-root objects in Fig. 5d–f. When the threshold on size was large, the post-processing ROC curves dropped steeply indicating that too many root objects were removed because the size of these root objects was smaller than the corresponding threshold. This experiment proved that eccentricity and size were useful attributes to separate roots from non-root objects.

3.3.4 Evaluating training features via root signature

An advantage of the MI-ACE algorithm (in addition the lack of parameters that need to be determined) was that a discriminative root signature was estimated. This signature helped to identify the unique root attributes that distinguish root pixels from soil. Figure 6 plots the estimated discriminative root target signatures. The signature bands of mean-R, mean-G, mean-B, mean-L, var-R, var-G and var-B had positive values with small variances, indicating that these features were generally larger in value for roots than for soil and background. The signature bands of H-a and H-b had negative values, indicating that roots generally had smaller values in these

features than background. The larger the absolute value was, the more informative that feature was in distinguishing roots from soil. The signature bands of mean-a, H-R, H-G, H-B, H-L, var-L, var-a and var-b were close to zeros which meant the difference between roots and soil at those feature bands was small. However, in some feature bands, the root signature values were widely spread. At a widely spread band, the signature value depended on the initial value and the differences between training results were large, for example, the value of signature at mean-b in Fig. 6.

3.3.5 Evaluating sensitivity of MIL methods to features

The features in these signatures in Fig. 6 can be divided into three groups by their corresponding signature value. The first group was the signature features with large average absolute signature value and small variance, the second group was the features that had small average absolute signature value with small variance, and the third group was features whose signature value had large variance. The comparison experiments with different selected features were done on the same training samples to examine the effects of features with different type of signature value. The experiments using all 18 features were compared with experiments using 9 and 17 features. The features which had large average absolute signature values with small variances were used in experiments with 9 features. Those feature were mean-R, mean-G, mean-B, mean-L, var-R, var-B, var-B, H-a and H-b. In the experiment with 17 features, the mean-b feature with very large variance was removed from the feature set and all other features were kept (Table 3).

For MI-ACE and miSVM methods, the features with large average absolute value were important for detection results as seen in Tables 2 and 3. It was also noted that the variances of detection results of MI-ACE and miSVM were significant improved by removing features which had large signature variance and small average absolute signature value. As seen in Tables 2 and 3, the variance of TPR with 9 features was one order of magnitude smaller than the variance of TPR with 18 features. Therefore, large average absolute signature values dominated the detection results and the larger the value was,

Table 2 Mean and variance of TPR of MI-ACE at FPR ranging from 0.01 to 0.06. The number of features of training dataset was varied to be 9, 17 or 18

FPR	MI-ACE-9		MI-ACE-17		MI-ACE-18	
	Mean	Var	Mean	Var	Mean	Var
0.01	**0.68**	**0.11e-4**	0.66	0.54e-3	0.66	0.38e-3
0.02	**0.74**	**0.03e-4**	0.72	0.34e-3	0.72	0.33e-3
0.03	**0.77**	**0.04e-4**	0.76	0.26e-3	0.75	0.32e-3
0.04	**0.79**	**0.03e-4**	0.78	0.21e-3	0.77	0.29e-3
0.05	**0.80**	**0.04e-4**	0.80	0.19e-3	0.79	0.25e-3
0.06	**0.81**	**0.04e-4**	0.81	0.16e-3	0.80	0.23e-3

Bold reflects the best performance

Table 3 Mean and variance of TPR of miSVM at FPR ranging from 0.01 to 0.06. The number of features of training dataset varied to be 9, 17 or 18

FPR	miSVM-9		miSVM-17		miSVM-18	
	Mean	Var	Mean	Var	Mean	Var
0.01	0.67	0.86e-4	0.69	1.84e-04	**0.70**	**2.10e-04**
0.02	**0.75**	**0.17e-4**	0.75	1.19e-04	0.75	1.17e-04
0.03	**0.78**	**0.10e-4**	0.78	0.88e-04	0.78	1.03e-04
0.04	**0.80**	**0.06e-4**	0.80	0.70e-04	0.80	0.82e-04
0.05	**0.81**	**0.05e-4**	0.81	0.55e-04	0.81	0.74e-04
0.06	**0.82**	**0.03e-4**	0.82	0.48e-04	0.82	0.64e-04

Bold reflects the best performance

Table 4 Mean and variance of TPR of MIForests at FPR ranging from 0.01 to 0.06. The number of features of training dataset varied to be 9, 17 or 18

FPR	MIForests-9		MIForests-17		MIForests-18	
	Mean	Var	Mean	Var	Mean	Var
0.01	0.36	1.3e-3	**0.51**	**2.0e-3**	0.49	3.7e-3
0.02	0.40	1.5e-3	**0.56**	**2.2e-3**	0.54	4.1e-3
0.03	0.43	1.5e-3	**0.60**	**2.3e-3**	0.58	4.1e-3
0.04	0.46	1.6e-3	**0.62**	**2.3e-3**	0.61	4.0e-3
0.05	0.48	1.7e-3	**0.65**	**2.3e-3**	0.63	3.8e-3
0.06	0.50	1.7e-3	**0.67**	**2.2e-3**	0.66	3.6e-3

Bold reflects the best performance

the more consistent the results were for the MI-ACE and miSVM methods.

In contrast, we found that the features with small variance were important for the MIForests method. The sensitivity analysis showed an improvement in the variance when the features with the largest variance were removed as shown in Table 4. Although the variance of detection using only the features with the small variance and large mean further improved the variance, it also reduced the TPR mean. Thus, the use of all features with small variance regardless of their mean values was important for the MIForests method.

3.3.6 Qualitative results

Root detection results were qualitatively compared as shown in Fig. 8 at threshold that FPR = 0.03 for different methods. The results were also post-processed by manually setting the thresholds of area and eccentricity to be 300 and 0.95, respectively. The thresholds were determined by identifying the threshold value that provided a boost in performance over results without post-processing. In general, the MIL methods outperformed the non-MIL methods. The miSVM method detected the most roots in images and performed the best among all these methods.

3.4 Experiment: effect of label accuracy

3.4.1 Sample versus image level label training

In this experiment, we compared the detection accuracy of the different methods when using training data paired with image level label versus training data paired with instance level label. The experimental runs using sample level label were trained using 10 images that were randomly selected from a total of 30 images that had been manually labeled. About 1000 samples labeled as root and 1000 samples labeled as soil were randomly selected in each image. In order to run MIL methods with instance level labeled data, we put only one instance per bag and the bag label was assigned to the label of the only instance in the bag.

ROC curves corresponding to instance and image level label accuracy can be seen in Fig. 7a–c. Among all three MIL methods, MI-ACE was the most robust algorithm to noisy labeled training data. This was indicated by the overlapping of the ROC curves regardless of label accuracy (Fig. 7a). MIForests and miSVM were more sensitive to label accuracy than MI-ACE (Fig. 7b, c). The ROC curves of MIForests and miSVM with image level label were lower than the ROC curves of MIForests and miSVM with instance level label. The results of MIForests and miSVM improved as the label became more accurate. The results also showed that the variance of ROC curves training with instance level data was larger than that training with image level data, especially for MI-ACE and miSVM. The results of training with image level data were more consistent because it was easy to get a large number of images with image level label, but it was expensive to label every instance and hard to get a large amount of training data with instance level label. A large amount of training data helped the generalization of the learning model so that the test results were more consistent.

3.4.2 Small bag versus image level label

We also compared results using training data paired with image level label to results using training data paired with
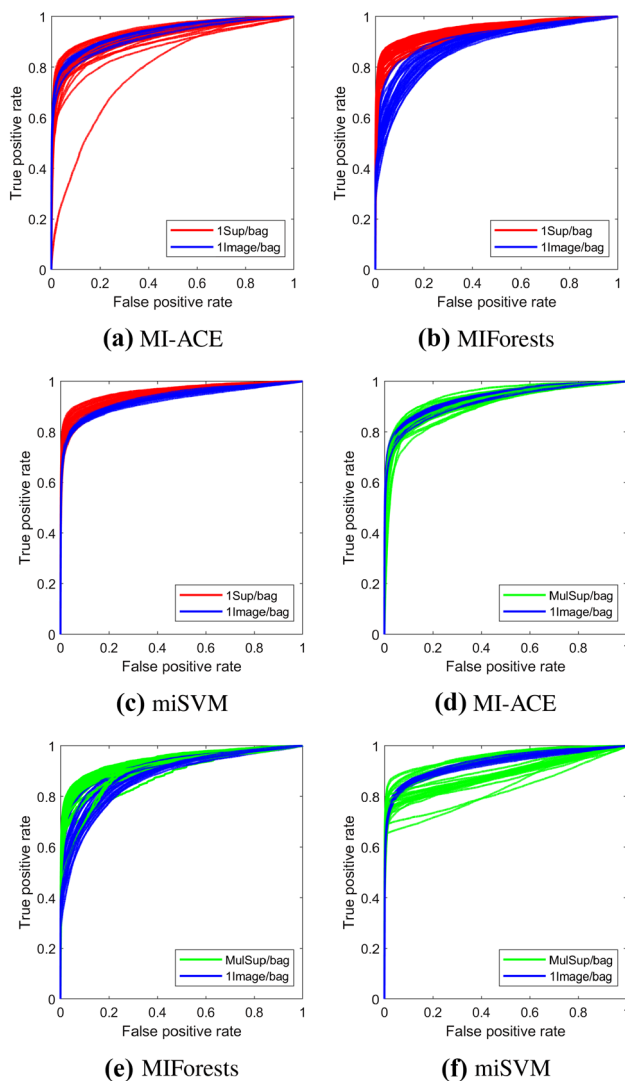
Fig. 7 ROC curves of MI-ACE, MIForests and miSVM with different levels of label accuracy of training data. Each algorithm is run 30 times

small bag level label (Fig. 8). Small bags were generated by SLIC algorithm which were large enough to group about 10 adjacent instances into a bag. It was a positive bag if there were root samples in the bag. Otherwise, it was a negative bag. Similarly to training data with sample level label, the experiment runs using small bag level label were trained using 10 images that were randomly selected from a total of 30 images. About 100 positive bags and 100 negative bags were randomly selected in each image.

ROC curves corresponding to small bag level label and image level label can be seen in Fig. 7d–f. The results of MI-ACE and miSVM using image level label training data were comparable with results using small bag level label because the ROC curves of MI-ACE and miSVM with different label accuracy overlapped with each other. But the results using image level labeled data were more consistent than the

results using small bag level labeled data because there were more labeled data with image level label that improved the generalization of the learning model, indicating the importance of a large amount of data to the learning results. For MIForests, the detection results with small bag level training data were much better than that with image level training data and proved again the importance of precise label of training data to the learning results of MIForests.

4 Discussion

We compared three different MIL classifiers with same training data. These methods varied in segmentation performance. The miSVM algorithm outperformed the other approaches with a higher average TPR at the same FPR and MI-ACE was a close second. These methods varied in initialization strategies, optimization and model refinement strategies and the feature space used. For example, during initialization both miSVM and MIForests initialize instance labels by assigning each instance the same label as the bag, whereas, for MI-ACE, a target signature is selected from a positive bag and used to initialize individual instance labels. During training, the methods also use a variety of strategies to refine and update labels. For example, miSVM uses the model estimated from the last iteration to predict labels for each instance and, then, retrain and refine the model. MIForests randomly draws labels for instances in bag with respect to the probability distribution of labels predicted using the model trained from last iteration and retrains with these sampled labels. MI-ACE uses yet another refinement strategy in that it always selects the instance most likely to be target from each positive bag and uses these to update the target model. Also, these three different MIL algorithms work in different feature spaces. The miSVM algorithm transfers the input features into a high-dimensional feature space through a kernel mapping, MI-ACE works in a whitened feature space, and MIForests works in the original input feature space. We believe that miSVM outperforms the others due to the high-dimensional kernel mapping and, thus, very nonlinear decision boundary. MI-ACE, however, provides the advantage that the learned target signature is interpretable and provides explainability to the user. Features played an important role in the performance of different classifiers. The mean, variance and entropy of each band in the RGB and LAB color spaces were used as 18-dimensional features to represent each superpixel. This feature vector was effective at representing the color information of a set of pixels in a superpixel. We mainly considered the color information of each superpixel because the superpixels generated by SLIC were similar in shape. However, in future work, other high level spatial features such as object shape may be evaluated to determine differences between roots and soil pixels. We used SLIC to generate
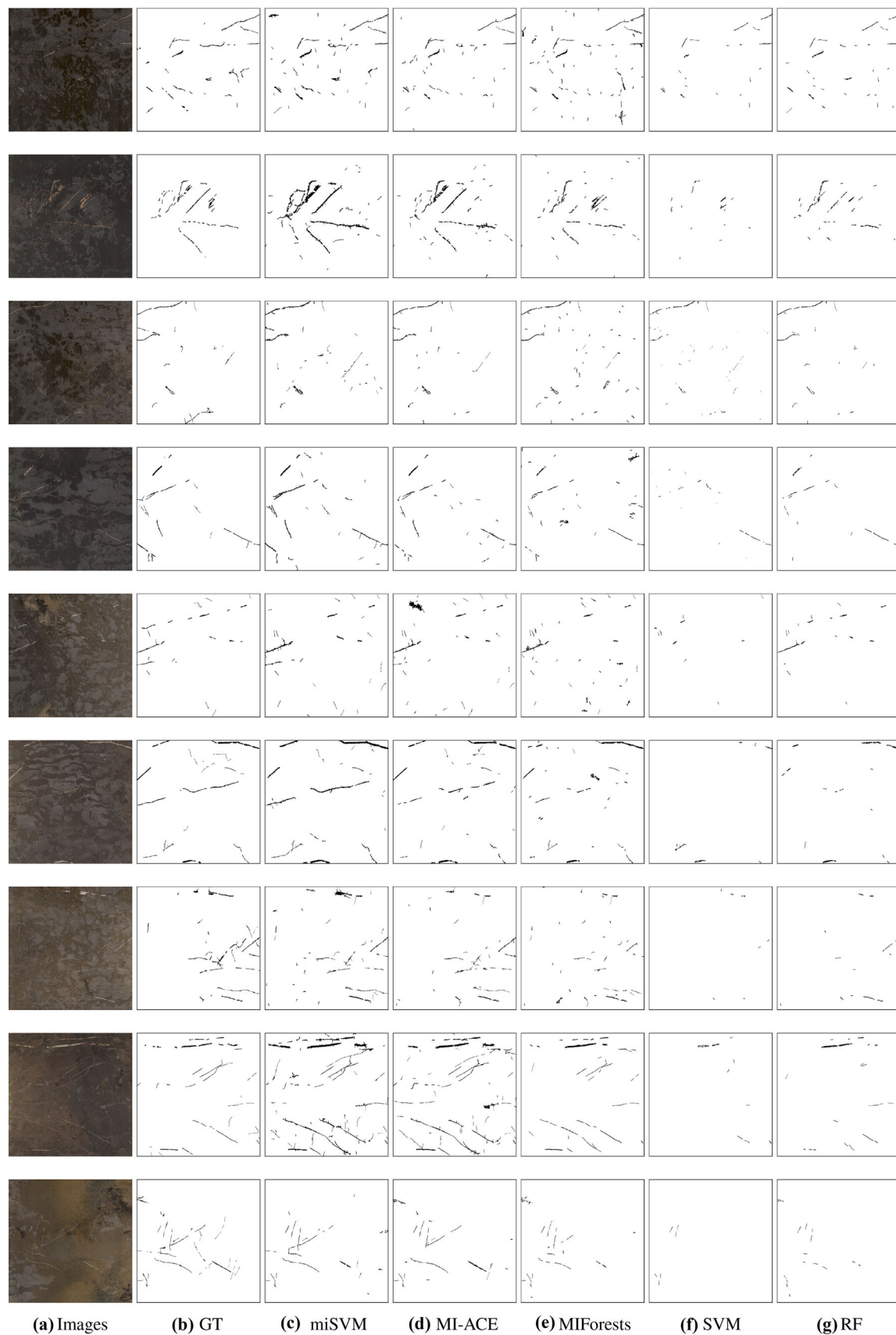
(a) Images   **(b)** GT   **(c)** miSVM   **(d)** MI-ACE   **(e)** MIForests   **(f)** SVM   **(g)** RF

Fig. 8 Qualitative examples of root segmentation results with different methods. The first column is the original images. The second column is the groundtruth (GT)

superpixels not only it can generate super pixels similar in size and shape but also it was the most feasible method to make superpixels contain pixels from same class via controlling the average size (or the totally number) of superpixels and the spatial regularization. Using SLIC method to generate superpixels avoided large variance in superpixel size that the performance deteriorated sharply by a misclassified large superpixel. Our ongoing work will extend the experiments to other recently available minirhizotron root datasets to do evaluation in a larger scale.

5 Summary

We propose the MIL approach to detect roots in minirhizotron images. Our MIL approach uses image-level labeled data to segment roots which significantly reduces the effort needed to label training data for the application of supervised learning algorithm. We tested the capability of three MIL methods to achieve competitive root detection over a range of soil textures and colors. Our superpixel feature histogram-based downsampling strategy can effectively increase the target ratio in positive bag that the image-level labeled training data achieve comparable results to training data labeled in finer scale. The miSVM method performed the best in detecting roots with image-level labels among all three MIL algorithms. Overall, our results suggest application of MIL methods can substantially improve the imaging analysis bottleneck currently occurring in the study of roots by the minirhizotron technique.

References

1. Liu, K., He, A., Ye, C., Liu, S., Lu, J., Gao, M., Fan, Y., Lu, B., Tian, X., Zhang, Y.: Root morphological traits and spatial distribution under different nitrogen treatments and their relationship with grain yield in super hybrid rice. Sci. Rep. **8**, 131 (2018)

2. Hammer, G.L., Dong, Z., McLean, G., Doherty, A., Messina, C., Schussler, J., Zinselmeier, C., Paszkiewicz, S., Cooper, M.: Can changes in canopy and/or root system architecture explain historical maize yield trends in the US corn belt? Crop. Sci. **49**, 299–312 (2009)

3. Taylor, H.M., Upchurch, D.R., Brown, J.M., Rogers, H.H.: Some methods of root investigations. In: McMichael, B.L., Persson, H. (eds.) Plants Roots and Their Environment, pp. 553–564. Elsevier Science Publisher B. V., New York (1991)

4. Bates, G.H.: A device for the observation of root growth in the soil. Nature **139**, 966–967 (1937)

5. Rewald, B., Ephrath, J.E.: Minirhizotron techniques. In: Eshel, A., Beeckman, T. (eds.) Plant Roots: The Hidden Half, 4th edn, pp. 421–429. CRC Press, Hoboken (2013)

6. Waddington, J.: Observation of plant roots in situ. Can. J. Bot. **49**, 1850–1852 (1971)

7. Johnson, M.G., Tingey, D.T., Phillips, D.L., Storm, M.J.: Advancing fine root research with minirhizotrons. Environ. Exp. Bot. **45**, 263–289 (2001)

8. Shojaedini, S.V., Heidari, M.: A new method for root detection in minirhizotron images: hypothesis testing based on entropy-based geometric level set decision. IJE Trans. A Basics **27**, 91–100 (2013)

9. Zeng, G., Birchfield, S.T., Wells, C.E.: Detecting and measuring fine roots in minirhizotron images using matched filtering and local entropy thresholding. Mach. Vis. Appl. **17**, 265–278 (2006)

10. Zeng, G., Birchfield, S.T., Wells, C.E.: Rapid automated detection of roots in minirhizotron images. Mach. Vis. Appl. **21**, 309–317 (2010)

11. Xu, W., Yu, G., Zare, A., Zurweller, B., Rowland, D., Reyes-Cabrera, J., Fritschi, F.B., Matamala, R., Juenger, T.E.: Overcoming Small Minirhizotron Datasets Using Transfer Learning. arXiv preprint arXiv:1903.09344 (2019)

12. Wang, T., Rostamza, M., Song, Z., Wang, L., McNickle, G., Iyer-Pascuzzi, A.S., Qiu, Z., Jin, J.: SegRoot: a high throughput segmentation method for root image analysis. Comput. Electron. Agric. **154**, 25–35 (2018)

13. Yasrab, R., Atkinson, J.A., Wells, D.M., French, A.P., Pridmore, T.P., Pound, M.P.: RootNav 2.0: Deep learning for automatic navigation of complex plant root architectures. GigaScience **8**(11), 123 (2019)

14. Smith, A.G., Petersen, J., Selvan, R., Rasmussen, C.R.: Segmentation of roots in soil with u-net. Plant Methods **16**(1), 1–15 (2020)

15. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: ECCV, pp. 483–499 (2016)

16. Rahmanzadeh, H., Shojaedini, S.V.: Novel automated method for minirhizotron image analysis: root detection using curvelet transform. IJE Trans. C Asp. **29**, 337–346 (2016)

17. Zare, A., Jiao, C., Glenn, T.: Discriminative multiple instance hyperspectral target characterization. IEEE Trans. Pattern Anal. Mach. Intell. **40**, 2342–2354 (2018)

18. Dietterich, T.G., Lathrop, R.H., Lozano-Perez, T.: Solving the multiple instance problem with axis-parallel rectangles. Artif. Intell. **89**, 31–71 (1997)

19. Bolton, J., Gader, P., Frigui, H., Torrione, P.: Random set framework for multiple instance learning. Inf. Sci. **181**, 2061–2070 (2011)

20. Chen, Y., Bi, J., Wang, J.Z.: MILES: Multiple-instance learning via embedded instance selection. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 1931–1947 (2006)

21. Maron, O., Lozano-Pèrez, T.: A framework for multiple-instance learning. In: Jordan, M.I., Kearns, M.J., Solla, S.A. (eds.) Advances in Neural Information Processing Systems 10, pp. 570–576. MIT Press, Cambridge (1998)

22. Zhang, Q., Goldman, S.: EM-DD: aan improved multiple-instance learning technique. In: Dietterich, T., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems 14, pp. 1073–1080. MIT Press, Cambridge (2002)

23. Zare, A., Gader, P.: Pattern recognition using functions of multiple instances. In: Proceedings of International Conference on Pattern Recognition, pp. 1092–1095 (2010)

24. Jiao, C., Zare, A.: Functions of multiple instances for learning target signatures. IEEE Trans. Geosci. Remote Sens. **53**, 4670–4686 (2015)

25. Shrivastava, A., Pillai, J. K., Patel, V. M., Chellappa, R.: Dictionary-based multiple instance learning. In: 2014 IEEE International Conference on Image Processing, pp. 160–164 (2014)

26. Shrivastava, A., Patel, V.M., Pillai, J.K., Chellappa, R.: Generalized dictionaries for multiple instance learning. Int. J. Comput. Vis. **114**, 288–305 (2015)

27. Leistner, C., Saffari, A., Bischof, H.: MIForests: multiple-instance learning with randomized trees. In: Daniilidis, K., Maragos, P.,

Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6316, pp. 29–42. Springer, Heidelberg (2010)

28. Andrews, S., Tsochandaridis, I., Hofman, T.: Support vector machines for multiple-instance learning. Adv. Neural. Inf. Process. Syst. **15**, 561–568 (2003)

29. Chang, Y., Yan, L.X., Wu, T., Zhong, S.: Remote sensing image stripe noise removal: from image decomposition perspective. IEEE Trans. Geosci. Remote Sens. **54**(12), 7018–7031 (2016)

30. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC Superpixels. EPFL Technical Report (2010)

31. Broadwater, J., Chellappa, R.: Hybrid detectors for subpixel targets. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 1891–1903 (2007)

32. Kraut, S., Scharf, L.L.: The CFAR adaptive subspace detector is a scale-invariant GLRT. IEEE Trans. Signal Process. **47**, 2538–2541 (1999)

33. Kraut, S., Scharf, L.L., McWhorter, L.T.: Adaptive subspace detectors. IEEE Trans. Signal Process. **49**, 1–16 (2001)

34. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**, 273–297 (1995)

35. Schmer, M.R., Vogel, K.P., Mitchell, R.B., Perrin, R.K.: Net energy of cellulosic ethanol from switchgrass. Proc. Natl. Acad. Sci. USA **105**, 464–469 (2008)

36. Comas, L.H., Becker, S.R., Cruz, V.M., Byrne, P.F., Dierig, D.A.: Root traits contributing to plant productivity under drought. Front. Plant. Sci. **4**, 442 (2013)

37. de Graaff, M.-A., Six, J., Jastrow, J.D., Schadt, C.W., Wullschleger, S.D.: Variation in root architecture among switchgrass cultivars impacts root decomposition rates. Soil Biol. Biochem. **58**, 198–206 (2013)

38. Berhongaray, G., Janssens, I.A., King, J.S., Ceulemans, R.: Fine root biomass and turnover of two fast-growing poplar genotypes in a short-rotation coppice culture. Plant Soil **373**, 269–283 (2013)

39. Powers, D.M.: Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. J. Mach. Learn. Technol. **2**, 37–63 (2011)

40. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011)

41. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)

42. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Guohao Yu received the B.S. degree in electrical engineering from the Harbin Engineering University, Harbin, China, and the M.S. degree in optical engineering from the Chinese Academy of Sciences,Shanghai, China in 2008 and 2011, respectively. He is currently a graduate research assistant working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Florida. His research interests include multiple instance learning, semantic segmentation and minirhizotron image analysis.

Alina Zare received the Ph.D. degree from the University of Florida, Gainesville, in 2008. She is currently an associate professor in the Department of Electrical and Computer Engineering, University of Florida. Her research interests include machine learning and pattern recognition, sparsity promotion, target detection, hyperspectral image analysis and remote sensing. She is a recipient of the 2014 National Science Foundation CAREER award and the 2014 National Geospatial-Intelligence Agency New Investigator Program Award. She is a senior member of the IEEE.

Hudanyun Sheng received the B.Sc. degree in physics from Tongji University, Shanghai, China, in 2015, the M.Sc. degree in Industrial and Systems Engineering from University of Florida, Gainesville, Florida. She is now working toward the M.Sc. degree in Electrical and Computer Engineering in University of Florida and at the same time working as a graduate research assistant in the Machine Learning and Sensing Lab.

Roser Matamala received her M.Sc. and Ph.D. degrees in biological sciences from the University of Barcelona, Spain, in 1993 and 1997, respectively. She was a Research Associate at Duke University in 1998 and then at Argonne National Laboratory in 2000. Currently, she is a Scientist at Argonne and a Fellow at the Northwestern-Argonne Institute of Science and Engineering and at the UChicago Consortium for Advanced Science and Engineering. She is an ecologist and biogeochemist who studies the carbon, water and energy cycles at the ecosystem level and the changes occurring due to human factors, such as land-use and global climate changes. She is an investigator for many Department of Energy, Office of Science, Biological and Environmental Research projects. In these projects, she has been mapping the stocks and characterizing the composition of soil organic carbon in the Arctic region. She is also studying climate adaptation and sustainability of bioenergy crops across continental scale environmental gradients. In addition, she is developing new sensors for detecting changes in roots and soil moisture funded by ARPA.E and NSF. She has published over 50 articles in scientific journals, book chapters and scientific reports.

Joel Reyes-Cabrera was born in Leon, Nicaragua, where he lived 17 years and then moved to Costa Rica. He received his B.Sc. degree in agricultural sciences from EARTH University at Costa Rica in 2009. He entered graduate school at the University of Florida and received his M.Sc. and Ph.D. degrees in 2013 and 2017, respectively. His main areas of research interests are agricultural water management, plant physiology, bioenergy and carbon fluxes in the agroecosystem. Dr. Reyes-Cabrera is currently a postdoctoral fellow in the Department of Plant Sciences at the University of Missouri where he conducts research on the root biology of bioenergy crops.

Felix B. Fritschi is a native of Peru and received an Ing. HTL degree from the Swiss College of Agriculture, an M.S. in Agronomy from the University of Florida and a Ph.D. in Plant Biology from the University of California, Davis. He joined the University of Missouri in 2007 and currently is the C. Alice Donaldson Professor of Bioenergy Crop Physiology and Genetics in the Division of Plant Sciences. Dr. Fritschi is a member of the Interdisciplinary Plant Group at the University of Missouri and a Fellow of the Crop Science Society of America. Dr. Fritschi has a long-standing interest in sustainable crop production and an active research program aimed at enhancing crop resource use efficiency, particularly as related to production under drought and heat stress conditions. Dr. Fritschi's program leverages physiological mechanism-based strategies and genetics to better understand and ultimately reduce the impacts of abiotic stress on crop plants. He is also engaged in the development of innovative tools and techniques to better characterize plant growth under field conditions, exploit natural genetic variation and accelerate crop improvement.

Thomas E. Juenger is a Professor in the Department of Integrative Biology at the University of Texas at Austin. He received his Ph.D. in Ecology and Evolutionary Biology at the University of Chicago (1999) and completed postdoctoral research at the University of California Berkeley as a Miller Fellow (1999–2002). Tom's research focuses on interface of ecological and evolutionary quantitative genetics in natural plant populations. Much of his research has focused on studies of the genetics of physiological traits, abiotic stress tolerance and local adaptation and gene-by-environment interaction. More recently, he has led several projects exploring local adaptation in C4 perennial grasses including studies of plant root systems.