Received: 27 April 2022 | Accepted: 6 June 2022

DOI: 10.1002/ppj2.20051

The Plant Phenome Journal OPEN ACCESS

ORIGINAL RESEARCH

Plant phenotyping with limited annotation: Doing more with less

Koushik Nagasubramanian[1] | **Asheesh Singh**[2] | **Arti Singh**[2] | **Soumik Sarkar**[3] | **Baskar Ganapathysubramanian**[3]

[1]Dep. of Electrical Engineering, Iowa State Univ., Ames, IA, USA

[2]Dep. of Agronomy, Iowa State Univ., Ames, IA, USA

[3]Dep. of Mechanical Engineering, Iowa State Univ., Ames, IA, USA

Correspondence
Baskar Ganapathysubramanian, Dep. of Mechanical Engineering, Iowa State Univ., Ames, IA, USA.
Email: baskarg@iastate.edu

Assigned to Associate Editor Michael Gore.

Funding Information
AI Institute for Resilient Agriculture (USDA-NIFA #2021-67021-35329), COALESCE: COntext Aware LEarning for Sustainable CybEr-Agricultural Systems (CPS Frontier # 1954556), FACT: A Scalable Cyber Ecosystem for Acquisition, Curation, and Analysis of Multispectral UAV Image Data (NIFA #2019-67021-29938), Smart Integrated Farm Network for Rural Agricultural Communities (SIRAC) (NSF S&CC #1952045), Plant Sciences Institute faculty fellowship (PSI), Iowa Soybean Association, and USDA CRIS Project IOW04714.

Abstract

Deep learning (DL) methods have transformed the way we extract plant traits—both under laboratory as well as field conditions. Evidence suggests that “well-trained” DL models can significantly simplify and accelerate trait extraction as well as expand the suite of extractable traits. Training a DL model typically requires the availability of copious amounts of annotated data; however, creating large-scale annotated dataset requires nontrivial efforts, time, and resources. This limitation has become a major bottleneck in deploying DL tools in practice. Self-supervised learning (SSL) methods give exciting solution to this problem, as these methods use unlabeled data to produce pretrained models for subsequent fine-tuning on labeled data and have demonstrated superior transfer learning performance on down-stream classification tasks. We investigated the application of SSL methods for plant stress classification using few labels. We select a plant stress classification problem to test the effectiveness of SSL, as it is a fundamentally challenging problem due to (a) disease classification which depends on the abnormalities in a small number of pixels, (b) high data imbalance across different classes, and (c) fewer annotated and available plant stress images than in other domains. We compared seven SSL approaches spanning four broad classes of SSL methods on soybean [*Glycine max* L. (Merr.)] plant stress dataset and report that pretraining on unlabeled plant stress images significantly outperforms transfer learning methods using random initialization for plant stress classification. In summary, SSL-based model initialization and data curation improves annotation efficiency for plant stress classification tasks and will circumvent data annotation challenges associated with DL methods.

1 | INTRODUCTION

Prior to machine learning (ML) integration, plant sciences were using manual phenotyping methods that were labor intensive and at times ineffective for object identification, classification, quantification, and prediction (A. Singh et al., 2016). Then with advances in drone (Feng et al., 2021; Guo et al., 2021), ground robots (Atefi et al., 2021; Gao et al.,

Abbreviations: BYOL, Bootstrap Your Own Latent; DL, deep learning; ML, machine learning; Moco, Momentum Contrast SSL algorithm; NNBYOL, Nearest-Neighbor Bootstrap Your Own Latent; SimCLR, Simple Framework for Contrastive Learning of Visual Representations; SSL, self-supervised representation learning; SwaV, Swapping Assignments between Views; VICReg, Variance-Invariance-Covariance Regularization; CNN, Convolutional Neural Networks.

2018; Riera et al., 2021), sensors (K. Parmley et al., 2019; Pieruschka & Schurr, 2019), high-throughput phenotyping and phenomics took over (Araus & Cairns, 2014; A. K. Singh et al., 2021). These advancements made the measurement activities of multiple plant traits at various growth stages rapid, precise and accurate. Machine learning, particularly deep learning (DL), techniques can be effectively used to make sense of all these collected data (Jiang & Li, 2020; K. A. Parmley et al., 2019). Supervised ML methods trained with a lot of labeled data were effectively able to solve complex phenotyping tasks (Akintayo et al., 2018; Jubery et al., 2021; Nagasubramanian et al., 2018; Pound et al., 2017), in particular plant stress phenotyping. Large-scale plant stress phenotyping has the potential to transform disease scouting, crop management and breeding for climate change (D. P. Singh et al., 2021). In recent years, there has been an increasing push in applying deep learning techniques for automating plant stress classification and quantification (A. Singh et al., 2021; A. K. Singh et al., 2018), and there has been advances in interpretability of these models (Nagasubramanian et al., 2020; J. Shook et al., 2021), as well as privacy preserving DL models for plant stress phenotyping (Cho et al., 2021).

Despite the interest in DL, one of the critical drawbacks of training a DL model successfully for large-scale plant stress phenotyping applications is the availability of copious amounts of carefully annotated data. Creating a large-scale annotated dataset requires nontrivial efforts, domain knowledge, time, and resources. To overcome this drawback of supervised learning models, one effective and practical strategy is to use transfer learning (Ghosal et al., 2018; Zhuang et al., 2021).

In transfer learning, a model is generally pretrained using an abundant amount of labeled training data to solve a task in the source domain and is then fine-tuned with a small(er) number of labels for solving the target domain task. Ideally, the source domain and target domain tasks should be closely related for the model to be able to be transferable. The data efficiency to solve the target task depends on the amount of similarity between the source and target domains (Zhuang et al., 2021). Because DL models generally require a large dataset of images to learn from scratch, using transfer learning to fine-tune the pretrained model architecture can significantly improve model performance on the target task. Traditionally, transfer learning has utilized supervised pretraining, where the model is trained in the source domain using a large amount of labeled data or with unsupervised models like autoencoders that rely on an image reconstruction objective (Weiss et al., 2016). The generalization of convolutional neural network-based autoencoders for transfer learning is domain dependent and these models might require large amount of labeled data for fine-tuning as they are not always annotation efficient (Bank et al., 2020). Weak-supervision (Bellocchio et al., 2019; Ghosal et al., 2019; Marino et al., 2019), semi-supervised (Khaki et al., 2021; Pérez-Ortiz et al., 2015), synthetic data creation (Giuffrida et al., 2017), and active learning (Chandra et al., 2020; Nagasubramanian, et al., 2021; Rawat et al., 2022) techniques have also been used to reduce the amount of labeling needed for plant phenotyping tasks. Active learning methods adaptively select the most informative samples for labeling for the highest improvement in test accuracy (Ren et al., 2022; Settles, 2009). A recent alternate has been to use large amounts of unlabeled data to pretrain a model, with an approach called self-supervised pretraining (Jing & Tian, 2021; Kar et al., 2021; Misra & van der Maaten, 2020). This allows using large amount of data from the target domain itself to pretrain a model using pretext tasks. Figure 1 shows the self-supervised pretraining pipeline used for soybean [*Glycine max* L. (Merr.)] stress classification.

Core Ideas

- Self-supervised learning (SSL)-based pretraining provides excellent model initializations.
- Self-supervised representations are annotation efficient and transferable for soybean stress classification.
- Barlow Twins was the best SSL method for annotation efficiency.

Self-supervised pretraining using unlabeled data has been successfully used in natural image recognition (using benchmark datasets like ImageNet and CIFAR-10) tasks especially when labeled examples are scarce (T. Chen, Kornblith, Swersky, et al., 2020). Self-supervised representation learning (SSL) aims to learn effective visual representations without any human supervision (T. Chen, Kornblith, Norouzi, et al., 2020). In other words, a good SSL strategy will produce a network that learns a good latent representation without any labels as (red box in Figure 1). This model (i.e., its weights) and associated latent representation can then be used for downstream tasks (bottom row in Figure 1). These methods have made significant progress over the last few years, almost reaching the performance of supervised learning models on many downstream transfer learning tasks like image classification and object detection. SSL methods have been successfully applied for medical image classification (Azizi et al., 2021), satellite remote sensing (Ayush et al., 2020) and biodiversity monitoring (Pantazis et al., 2021). Recently, SSL methods have been applied in agriculture domain for phenotyping tasks in seeds (Margapuri & Neilsen, 2021), irrigated fields (Agastya et al., 2021), weeds (Güldenring & Nalpantidis, 2021), and crops (Herrera-Gerena et al., 2021; Marin Zapata et al., 2021; Nagasubramanian, et al., 2021).

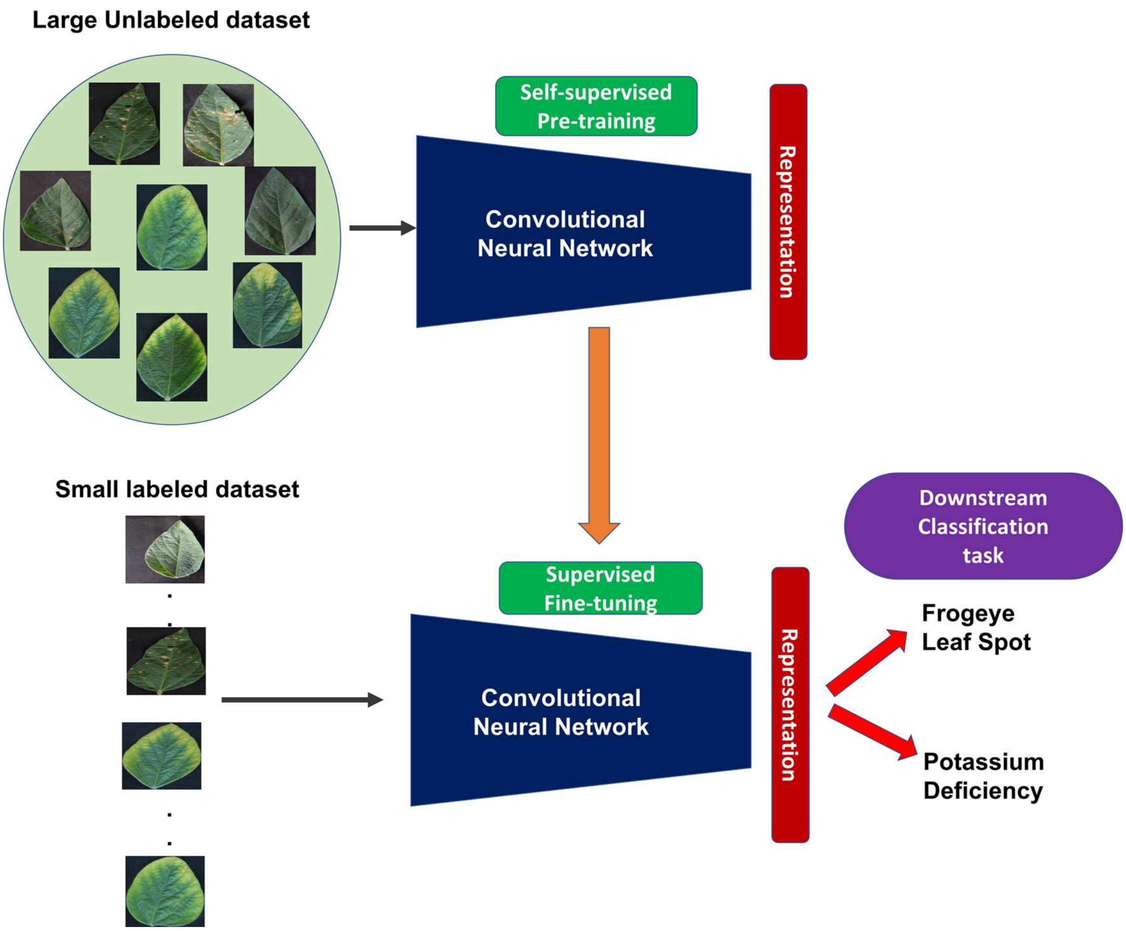

FIGURE 1 Illustration of self-supervised transfer learning using a hypothetical example for soybean abiotic and biotic stresses. In the first step, large amounts of unlabeled dataset are used for the self-supervised pretraining task, where the goal is to learn a good latent representation. Next, a small amount of labeled data is used to fine-tune this model, where the learnt representation is used for classification

Many recent SSL approaches rely on a joint embedding architecture in which two encoders are trained to produce similar embeddings for different views of the same image (Bardes et al., 2021). Although this is a rapidly evolving field, SSL approaches can be broadly classified into four different classes based on how the latent representation is learnt: contrastive learning (X. Chen, et al., 2020; Misra & van der Maaten, 2020), clustering (Caron et al., 2020), distillation (Chen & He, 2020; Grill et al., 2020), and information maximization (Bardes et al., 2021; Zbontar et al., 2021).

Plant stress classification is a fundamentally challenging problem because (a) stress type classification may depend on abnormalities in a small number of pixels (A. Singh et al., 2021), (b) of high data imbalance across different classes, and (c) there are fewer annotated plant stress images available than in other domains. In this work, we explored recent advances in self-supervised representation learning applied to plant stress classification using a very small number of labeled data. In particular, we compared seven different types of self-supervised pretraining methods (two using contrastive learning, one using clustering, two using distillation, and two using information maximization) methods for soybean stress classification. We report that self-supervised pretraining on the unlabeled data significantly improves the data curation process and annotation efficiency for image-based plant stress classification tasks. The use of self-supervised pretraining can remove the data annotation bottleneck in DL models.

2 | MATERIALS AND METHODS

2.1 | Dataset

The dataset consists of 16,573 RGB images of soybean leaves across nine different classes (i.e., eight different soybean stresses, and a class containing healthy soybean leaf). Figure 2 illustrates the nine different soybean leaf classes used in this study. Each stress class consisted of images from different cultivars at different growth stages. Images were taken from a diverse collection of accessions that were a part of genome wide association studies and field experiments of elite lines. More details on the dataset and imaging protocol can be found in (Ghosal et al., 2018). Briefly, these classes cover a diverse

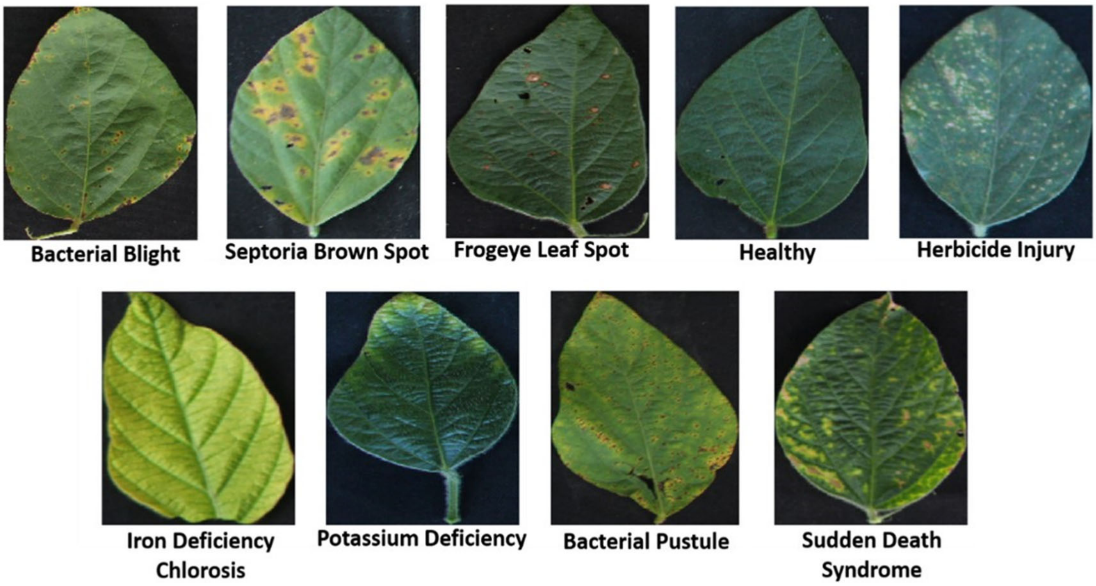

FIGURE 2 Illustration of stress symptoms on soybean leaflets across nine different classes (eight different soybean stresses and healthy soybean leaflets). These classes cover a diverse spectrum of biotic and abiotic foliar stresses

spectrum of biotic and abiotic stresses in soybean, with distinct foliar symptoms. The entire data set of 16,573 images consisted of class 0 = bacterial blight (1,524 images) [caused by *Pseudomonas syringae* pv. *glycinea*], Class 1 = septoria brown spot (1,358 images) [caused by *Septoria glycines*], Class 2 = frogeye leaf spot (1,122 images) [caused by *Cercospora sojina*], Class 3 = healthy (4,223 images), Class 4 = herbicide injury (1,395 images), Class 5 = iron deficiency chlorosis (1,844 images), Class 6 = potassium deficiency (2,186 images), Class 7 = bacterial pustule (1,674 images) [caused by *Xanthomonas axonopodis* pv. *glycines*], and Class 8 = sudden death syndrome (1,247 images) [caused by *Fusarium virguliforme*].

2.2 | Methods

As stated earlier, SSL techniques may be broadly classified into four classes, based on the approach to construct (or rather, constrain) the latent representations. In contrastive learning, the representations of an image and its augmented version (scaled, cropped, rotated, etc.) are made similar, whereas the representations of an image and other images different content are made dissimilar. Clustering-based methods aims to group similar set of images into same cluster. These methods group the latent representation space into different clusters without using labels. Distillation-based methods are also trained by maximizing the similarity of representations between an image and its transformed version. Unlike contrastive learning methods, these methods do not need negative examples (dissimilar images) for training. Information maximization methods maximize the information content in each dimension of the embeddings and minimizes the correlation of information across the dimensions. In this section, we briefly describe the seven different self-supervised pretraining methods evaluated for soybean stress classification in this work.

2.2.1 | Contrastive Learning (SimCLR and Moco-v2)

SimCLR (Simple Framework for Contrastive Learning of Visual Representations) is a contrastive learning-based SSL pretraining method (T. Chen, Kornblith, Norouzi, et al., 2020). It learns representations by maximizing agreement between differently transformed views of the same image and minimizing agreement between transformed views of different images via a contrastive loss. It considers different transformed views of the same image as positive pairs whereas all the other images in a batch are considered as negative pairs. This contrastive loss causes the representations of the positive pairs to attract each other, whereas representations of negative pairs repel each other.

Moco-v2 is a contrastive learning-based SSL pretraining method that learns representation by maximizing the agreement between different augmented views of the same image and minimizing the agreement between different images. It is an improved version of the Momentum Contrast SSL algorithm (Moco) (X. Chen, et al., 2020). It replaces the last fully connected layer in Moco with a two-layer multilayer perceptron head and ReLU activation for the pretraining. It uses two different encoder networks: one for query embedding and another for key embedding. It utilizes a memory bank to store embeddings from previous batches and use them as negative pairs. It also uses more data augmentation methods than Moco and needs fewer compute requirements than SimCLR. The query and keys are matched if they are encoded views of the same image.

2.2.2 | Clustering (SwaV)

Swapping Assignments between Views (SwaV) algorithm is an online clustering-based SSL method (Caron et al., 2020). This method groups different augmented views of an image into the same cluster. It learns representations by backpropagating gradients in a batch-wise manner. It uses a "swapped" prediction mechanism where cluster assignment of an image is predicted from the representation of augmented view of the same image. This method uses the Sinkhorn-Knopp (Knight, 2008) algorithm for equipartition constraint and soft clustering assignment to avoid trivial solutions.

2.2.3 | Distillation (BYOL and NNBYOL)

Bootstrap Your Own Latent (BYOL) is a distillation-based SSL method (Grill et al., 2020). It uses two neural networks as encoders to learn representations. The two encoder networks used in BYOL are referred to as online and target encoder. Both encoder networks have the same architecture. It uses an asymmetric learning rule to interact and learn from each network. Unlike contrastive learning methods, it can learn representations from raw image data without using negative pairs. Nearest-Neighbor Bootstrap Your Own Latent (NNBYOL) utilizes nearest neighbors from the latent space as positive examples to increase semantic variations during training and uses the same learning mechanism as BYOL.

2.2.4 | Information maximization (Barlow Twins and VICReg)

Barlow Twins is an information maximization-based SSL method (Zbontar et al., 2021). It learns feature representations using invariance and redundancy reduction-based loss functions. The model is trained by measuring the cross-correlation matrix between the outputs of two identical neural networks fed with different augmented versions of a same image and making them as close as possible to the identity matrix. The training attempts to zero out all the off-diagonal elements of the cross-correlation matrix. This maximizes the information content in the embeddings.

Variance-Invariance-Covariance Regularization (VICReg) (Bardes et al., 2021) is a method that tries to maximize the information content of the embeddings like Barlow twins. It uses a loss function with three components: invariance, variance, and covariance. The square distance between embedding vectors is used for the invariance loss component. It maintains the standard deviation of each dimension of the representations over a batch using a hinge loss. It also uses a covariance loss to avoid the informational collapse in representations due to redundancy between embedding dimensions.

2.2.5 | Pretraining setup

We used 90% of the soybean dataset (i.e, 14,916 unlabeled images) for pretraining and 10% (i.e, 1,657 images) for testing. We compared four different types of self-supervised pretraining (contrastive learning, clustering, distillation, and information maximization) methods for the soybean stress classification. The ResNet18 backbone was pretrained with an initial learning rate of 0.3 and a cosine learning rate scheduler. We trained the four different SSL methods for 1,000 epochs. The SSL model checkpoint with the lowest loss on training data was saved and used for down-stream classification task. Figure 3 shows the self-supervised pretraining methods used for soybean stress classification.

2.2.5.1 Linear Probing

To evaluate the transfer of representations, a popular evaluation protocol is to freeze the backbone model and train a linear classifier on the final layer representation (Kolesnikov et al., 2019). This method is used to understand the effectiveness of SSL representations for down-stream classification. Here, we froze the ResNet18 backbone model and used the 512-dimensional representation from the final layer of the model to train a linear classifier. A linear classifier with 512 nodes was used for the soybean stress classification. We used different label fractions of training sets (1, 3, 5, 7, 10, 30, 50, 70, and 100%) for the classifier as shown in Figure 5. All the linear probing experiments were repeated three times. The labeled fraction was randomly sampled for each of these repetitions. The linear classifier was evaluated on the test dataset (1,657) images. The supervised learning model trained from scratch and linear classifier trained on representations derived from ImageNet initialized model serves as our baseline methods. We used a batch-size of 64 for label fractions less than 10% of training data, that is, 1, 3, 5, 7, and 10%; and a batch-size of 512 for other fractions of training data that were tested. The classifier was trained for 30 epochs with a learning rate of 1. See Figure 4a for the broad architecture of linear probing evaluation.

2.2.5.2 End-to-end fine-tuning evaluation

To evaluate the SSL model initializations, we fine-tuned the model end-to-end using supervised learning. We used different label fractions of training sets (1, 3, 5, 7, and 10%)

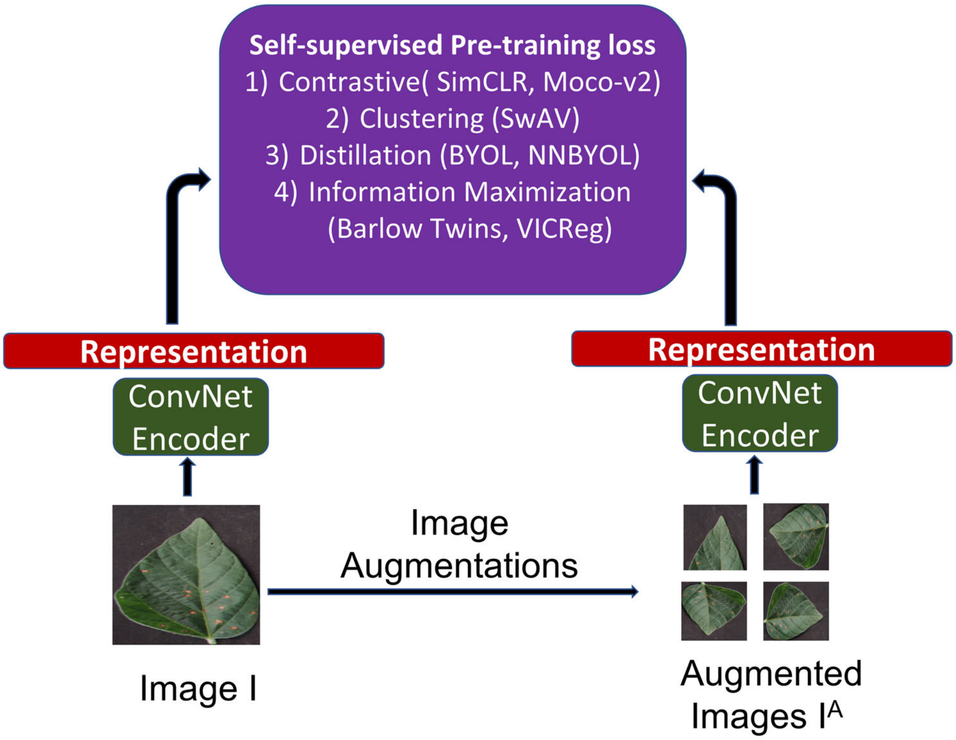

FIGURE 3 Overview of the four different classes of self-supervised pretraining methods applied on unlabeled soybean stress images. These four self-supervised learning (SSL) approaches compares the representations between image and its augmented versions using different loss functions. The model weights and latent representation from the pretrained models were used for the downstream soybean stress classification task using small amount of labeled data

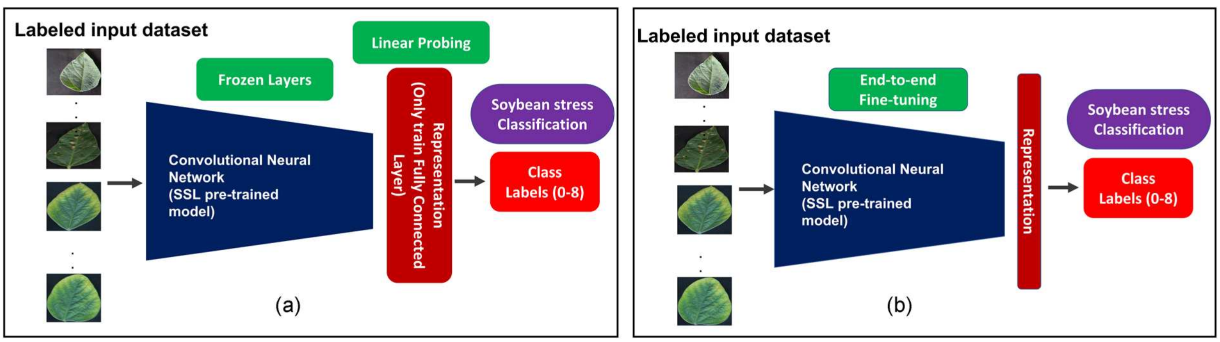

FIGURE 4 Illustration of (a) linear classification and (b) end-to-end fine-tuning methods, which were used to compare the accuracy of self-supervised learning (SSL) methods. We evaluated the representations from SSL pretraining methods using linear classification in Panel a) and evaluated the model initializations from SSL pretraining methods using linear classification in b). In Panel a, only weights of the last fully connected layer are fine-tuned, and in Panel b, all model weights are fine-tuned in the end-to-end evaluation

for fine-tuning the classifier as shown in Figure 6. Unlike the prior section, here we focus on accessing performance when there is a limited budget for labeling (set to 10% of the dataset). The classifier was evaluated on the test dataset (1,657) images. We used a batch-size of 32, learning rate of 0.001, and 50 number of epochs for training the classifier. The supervised learning from random initialization and ImageNet initialization serves as our baseline methods. All the end-to-end fine-tuning experiments were repeated three times. See Figure 4b for the broad architecture of end-to-end fine-tuning evaluation.

2.2.6 | Classification Accuracy

We calculate the multi-class classification accuracy from the confusion matrix: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TP and TN are

The Plant Phenome Journal OPEN ACCESS

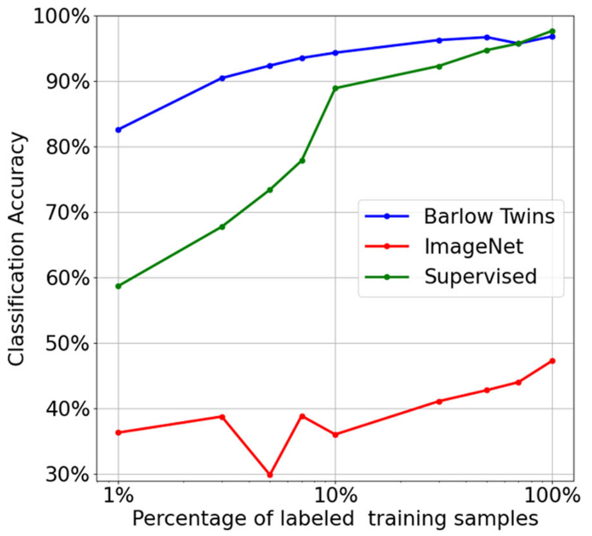

FIGURE 5 Linear evaluation of Barlow Twins method trained on different amounts of labeled images from soybean stress dataset. In this semi-log plot, the "Supervised" curve corresponds to training from scratch and the "ImageNet" curve represents a linear classifier trained on off-the-shelf features from ImageNet initialized model

the samples that were correctly classified by the model and are shown on the main diagonal of the confusion matrix. FP and FN are the samples that were incorrectly classified by the model. From these values, the classification accuracy is calculated as shown in Equation 1.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

3 | RESULTS AND DISCUSSION

3.1 | Linear probing

Barlow Twins was the most annotation efficient method for linear evaluation as shown in Figure 5. It outperformed the "supervised" model with random weights initialization when the labeled fraction is small. It also performed significantly better than off-the-shelf ImageNet features for linear evaluation. With just 10% of labeled samples, Barlow Twins obtained a mean classification accuracy (over three repetitions) of 95.49% using linear evaluation.

3.2 | Fine-tuning evaluation

Figure 6 shows performance of both the linear probing and end-to-end finetuning of Barlow Twins. For 3% labeled training samples, the Barlow Twins method obtained 93.03% mean classification accuracy (over three repetitions) for end-to-end fine-tuning and 90.44% mean classification accuracy (over three repetitions) for linear probing-based evaluation. It performed better than ImageNet initialization for end-to-end fine-tuning when the labeled fractions was less than 10% as shown in Figure 6b. Interestingly, our experiments indicate that at more than 10% of labeled data, both ImageNet initialization as well as SSL initialization perform in statistically similar ways. We attribute this to the smaller amount of domain-specific unlabeled image data used for SSL pretraining. Evidence suggests that performance of SSL models increases with the availability of larger unlabeled datasets (Cole et al., 2021).

With just 3% of labeled samples (447 images), Barlow Twins (and NNBYOL methods, as shown in later sections) were 10× more annotation efficient than active learning methods (trained from random initialization) (Nagasubramanian, et al., 2021). The benefit of active learning has also been shown to be negligible compared with SSL-based pretraining on benchmark datasets like CIFAR10 and CIFAR100 (Chan et al., 2021). The SSL Pretrained models significantly outperformed the supervised model trained from scratch for both linear and full-model fine-tuning.

Figure 7 shows the normalized confusion matrices for end-to-end evaluation of Supervised, ImageNet, and Barlow Twins methods. We can see that Barlow Twins was able to perform classification of soybean stress classes with small amounts of labeling data compared with other initialization methods. Barlow Twins outperformed ImageNet method for difficult to discriminate diseases like Bacterial blight (class 0) and bacterial pustule (class 7). These classes cause rating challenges for even expert raters during manual annotation due to similarity of disease symptoms (Nagasubramanian et al., 2020). The number of true positives increases for these two confounding classes as we move from supervised to ImageNet to self-supervised-based method.

3.3 | Comparison of Barlow Twins with other SSL methods

Barlow Twins and NNBYOL were the most annotation efficient methods for linear evaluation. All the SSL methods deployed outperformed the "supervised" model, especially when the labeled fraction of data is small. They also performed significantly better than off-the-shelf ImageNet features for linear evaluation as shown in Figure 8a. All the SSL pretraining methods outperformed supervised baseline for linear and end-to-end fine-tuning evaluation as shown in Figure 8a and b. For training label fractions greater than 5%, the performance of ImageNet initialization was on-par with SSL methods for the end-to-end fine-tuning. The performance of NNBYOL was comparable to Barlow Twins for both types of evaluation. Among the SSL methods, VICReg had the lowest performance.

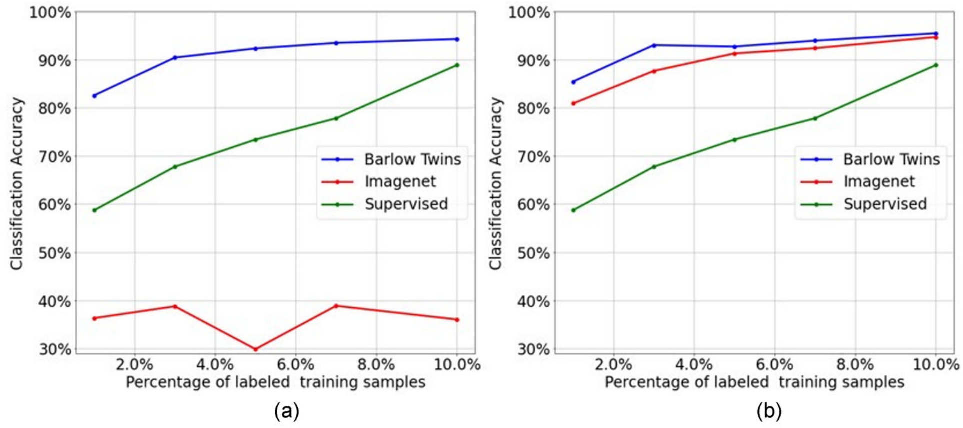

FIGURE 6 (a) Linear evaluation of Barlow Twins; (b) end-to-end fine-tuning evaluation of Barlow Twins. The supervised curve corresponds to training from scratch and the ImageNet curve represents a linear classifier trained on off-the-shelf features from ImageNet initialized model

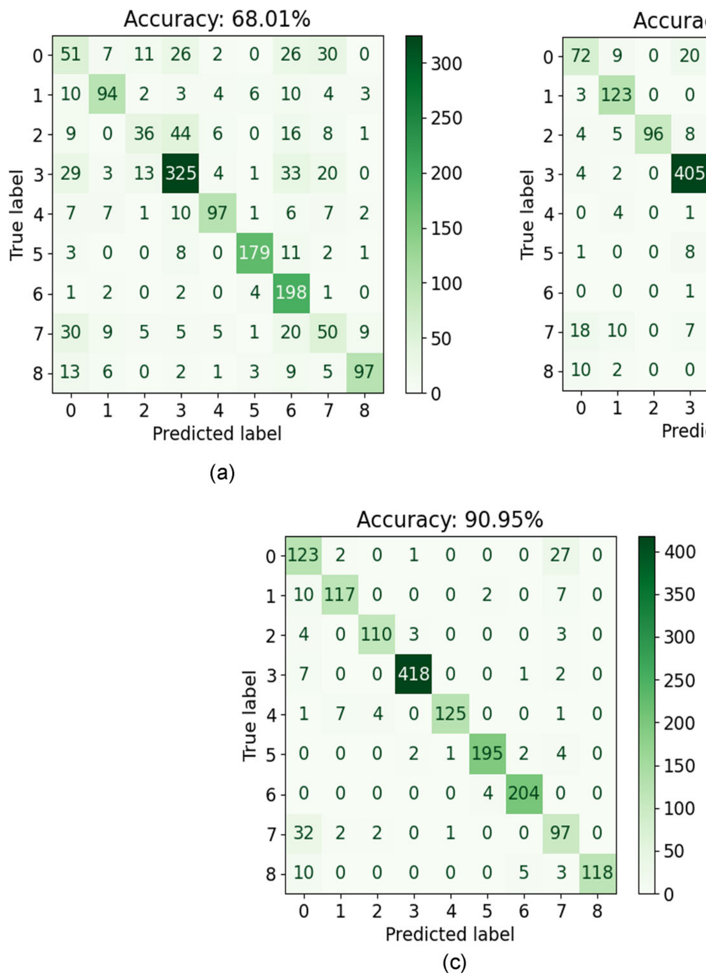

FIGURE 7 Normalized confusion matrices of test data (1,657 images) for (a) supervised, (b) ImageNet, and (c) Barlow Twins using 3% labeled training data. The nine classes are as following: 0 = bacterial blight, 1 = eptoria brown spot, 2 = frogeye leaf spot, 3 = healthy, 4 = herbicide injury, 5 = iron deficiency chlorosis, 6 = potassium deficiency, 7 = bacterial pustule, 8 = sudden death syndrome

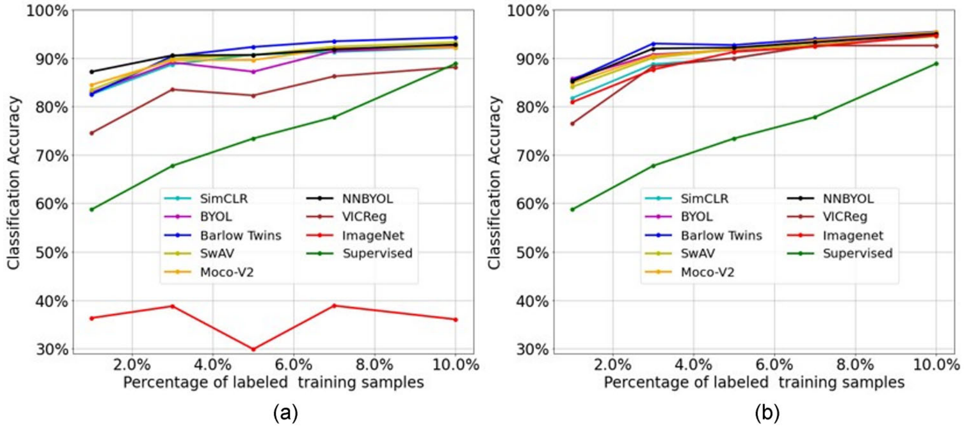

FIGURE 8 (a) Linear evaluation of self-supervised pretraining methods trained on different amounts of labeled images All the SSL methods performs significantly better than "Supervised" model when the labeled fraction is small. (b) End-to-end fine-tuning evaluation of self-supervised pretraining methods trained on different amounts of labeled data. The self-supervised learning (SSL) methods perform better than "ImageNet" when the labeled fraction is very small

The power of SSL approaches is magnified in the plant science domain particularly in crop production, plant breeding, and other related disciplines where high throughput phenotyping can produce massive amounts of unlabeled image data. We note that SSL can be successfully deployed across (a) spatial resolution (ground vs drone imaging vs satellite), (b) spectral scales (RGB vs multispectral vs hyperspectral), as well as (c) temporal data. The advantage of SSL was evident in the soybean plant stress dataset that we tested in our work. Although SSL has significant advantages, particularly in number of annotations needed for labelled data, there are open challenges that remain to be explored. For example, continual learning of SSL pretrained models with new data is an active research area (Purushwalkam et al., 2022). Another challenge with SSL is selecting the right combination of data augmentations for efficient pretraining in different domains. The right set of data augmentation methods must be selected based on the specific requirements (color or shape-based task) of the problem (Cole et al., 2021; Tian et al., 2020). For instance, we avoided color jittering-based data augmentation during SSL pretraining with soybean disease images to avoid drop in performance. Despite some of these challenges with SSL, we believe there are exciting avenues for the application of SSL such as in crop scouting and plant stress evaluation in an automated manner with drone or smartphones as developers and researchers can work with a smaller amount of labeled data for identification of classes. Although we do not demonstrate the assessment of level of severity for each image per class, this has been shown previously for soybean stress (Nagasubramanian et al., 2019; Naik et al., 2017). These applications can also include meta-genetic studies to leverage image data with genomic data (J. M. Shook et al., 2021).

4 | CONCLUSIONS

We found that SSL-based pretraining provides excellent model initializations and representations that are annotation efficient and transferable for soybean stress classification. Barlow Twins was the best SSL method for annotation efficiency. All the seven different SSL methods performed significantly better than the supervised baseline models for both linear and end-to-end evaluation. We observe that SSL methods were robust to class imbalances and highly efficient for soybean stress classification. The best SSL pretraining method for the phenotyping problem could be identified using the performance of the models in low data regime (less than 5% labeled training data). In addition, SLL-based methods are able to differentiate confounding stress classes with a smaller number of training images than supervised learning methods with ImageNet initialization.

There are several avenues of extending SSL approaches for plant phenotyping. Future works could consider (a) designing pretext tasks specific to plant phenotyping problems, (b) combining different SSL loss functions, (c) updating pretrained SSL models with unlabeled data from new classes, and (d) develop new SSL-based foundational models for annotation efficient image classification, segmentation, and object detection applications.

DATA AVAILABILITY STATEMENT

The data and code are available at https://github.com/koushik-n/SSL_soy.

ACKNOWLEDGMENTS

This work was supported by AI Institute for Resilient Agriculture (USDA-NIFA #2021-67021-35329), COALESCE:

COntext Aware LEarning for Sustainable CybEr-Agricultural Systems (CPS Frontier # 1954556), FACT: A Scalable Cyber Ecosystem for Acquisition, Curation, and Analysis of Multi-spectral UAV Image Data (NIFA #2019-67021-29938), Smart Integrated Farm Network for Rural Agricultural Communities (SIRAC) (NSF S&CC #1952045), Plant Sciences Institute faculty fellowship (PSI), Iowa Soybean Association, and USDA CRIS Project IOW04714. We thank team members of A. K. Singh and A. Singh groups for help in imaging and data collection.

Open access funding provided by the Iowa State University Library.

AUTHOR CONTRIBUTIONS

Koushik Nagasubramanian: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; Writing – review & editing. Asheesh Singh: Conceptualization; Data curation; Funding acquisition; Project administration; Resources; Supervision; Validation; Writing – original draft; Writing – review & editing. Arti Singh: Conceptualization; Data curation; Funding acquisition; Project administration; Resources; Supervision; Validation; Writing – original draft; Writing – review & editing. Soumik Sarkar: Conceptualization; Formal analysis; Funding acquisition; Project administration; Resources; Supervision; Writing – original draft; Writing – review & editing. Baskar Ganapathysubramanian: Conceptualization; Formal analysis; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing – original draft; Writing – review & editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Koushik Nagasubramanian https://orcid.org/0000-0003-2708-3945

Arti Singh https://orcid.org/0000-0001-6191-9238

REFERENCES

Agastya, C., Ghebremusse, S., Anderson, I., Reed, C., Vahabi, H., & Todeschini, A. (2021). Self-supervised contrastive learning for irrigation detection in satellite imagery. *arXiv*. https://arxiv.org/abs/2108.05484

Akintayo, A., Tylka, G. L., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). A deep learning framework to discern and count microscopic nematode eggs. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-27272-w

Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science*, *19*(1), 52–61. https://doi.org/10.1016/j.tplants.2013.09.008

Atefi, A., Ge, Y., Pitla, S., & Schnable, J. (2021). Robotic technologies for high-throughput plant phenotyping: Contemporary reviews and future perspectives. *Frontiers in Plant Science*, *12*, 611940. https://doi.org/10.3389/fpls.2021.611940

Ayush, K., Uzkent, B., Meng, C., Tanmay, K., Burke, M., Lobell, D., & Ermon, S. (2020). Geography-aware self-supervised learning. *arXiv*. https://arxiv.org/abs/2011.09980

Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., Loh, A., Karthikesalingam, A., Kornblith, S., Chen, T., Natarajan, V., & Norouzi, M. (2021). Big self-supervised models advance medical image classification. *arXiv*. https://arxiv.org/abs/2101.05224

Bank, D., Koenigstein, N., & Giryes, R. (2020). Autoencoders. *arXiv*. https://arxiv.org/abs/2003.05991

Bardes, A., Ponce, J., & LeCun, Y. (2021). VICReg: Variance-invariance-covariance regularization for self-supervised learning. *arXiv*. https://arxiv.org/abs/2105.04906

Bellocchio, E., Ciarfuglia, T. A., Costante, G., & Valigi, P. (2019). Weakly supervised fruit counting for yield estimation using spatial consistency. *IEEE Robotics and Automation Letters*, *4*(3), 2348–2355. https://doi.org/10.1109/LRA.2019.2903260

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised learning of visual features by contrasting cluster assignments. In *arXiv*. https://arxiv.org/abs/2006.09882

Chan, Y.-C., Li, M., & Oymak, S. (2021). *On the marginal benefit of active learning: Does self-supervision eat its cake?* Paper presented at the ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, Canada. https://doi.org/10.1109/ICASSP39728.2021.9414665

Chandra, A. L., Desai, S. V., Balasubramanian, V. N., Ninomiya, S., & Guo, W. (2020). Active learning with point supervision for cost-effective panicle detection in cereal crops. *Plant Methods*, *16*(1), 34. https://doi.org/10.1186/s13007-020-00575-8

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *arXiv*. http://arxiv.org/abs/2002.05709

Chen, T., Kornblith, S., Swersky, K., Norouzi, M., & Hinton, G. (2020). Big self-supervised models are strong semi-supervised learners. *arXiv*. https://arxiv.org/abs/2006.10029

Chen, X., Fan, H., Girshick, R., & He, K. (2020). Improved baselines with momentum contrastive learning. *arXiv*. https://arxiv.org/abs/2003.04297

Chen, X., & He, K. (2020). Exploring simple Siamese representation learning. *arXiv*. https://arxiv.org/abs/2011.10566

Cho, M., Nagasubramanian, K., Singh, A. K., Singh, A., Ganapathysubramanian, B., Sarkar, S., & Hegde, C. (2021). Privacy-preserving deep models for plant stress phenotyping. https://openreview.net/pdf?id=8JkuZSH487G

Cole, E., Yang, X., Wilber, K., Aodha, O. M., & Belongie, S. (2021). When does contrastive visual representation learning work? *arXiv*. https://arxiv.org/abs/2105.05837

Feng, L., Chen, S., Zhang, C., Zhang, Y., & He, Y. (2021). A comprehensive review on recent applications of unmanned aerial vehicle remote sensing with various sensors for high-throughput plant phenotyping. *Computers and Electronics in Agriculture*, *182*(106033), 106033. https://doi.org/10.1016/j.compag.2021.106033

Gao, T., Emadi, H., Saha, H., Zhang, J., Lofquist, A., Singh, A., Ganapathysubramanian, B., Sarkar, S., Singh, A., & Bhattacharya, S. (2018). A novel multirobot system for plant phenotyping. *Robotics*, *7*(4), 61. https://doi.org/10.3390/robotics7040061

Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(18), 4613–4618. https://doi.org/10.1073/pnas.1716999115

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., Singh, A. K., Singh, A., Hirafuji, M., Ninomiya, S., Ganapathysubramanian, B., Sarkar, S., & Guo, W. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics*, *2019*, 1525874.

Giuffrida, M. V., Scharr, H., & Tsaftaris, S. A. (2017, October). *ARIGAN: Synthetic Arabidopsis plants using generative adversarial network*. Paper presented at the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy. https://doi.org/10.1109/iccvw.2017.242

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Pires, B. A., Guo, Z. D., Azar, M. G., Piot, B., Kavukcuoglu, K., Munos, R., & Valko, M. (2020). Bootstrap your own latent: A new approach to self-supervised Learning. *arXiv*. https://arxiv.org/abs/2006.07733

Güldenring, R., & Nalpantidis, L. (2021). Self-supervised contrastive learning on agricultural images. *Computers and Electronics in Agriculture*, *191*(106510), 106510. https://doi.org/10.1016/j.compag.2021.106510

Guo, W., Carroll, M. E., Singh, A., Swetnam, T. L., Merchant, N., Sarkar, S., Singh, A. K., & Ganapathysubramanian, B. (2021). UAS-based plant phenotyping for research and breeding applications. *Plant Phenomics*, *2021*, 9840192. https://doi.org/10.34133/2021/9840192

Herrera-Gerena, J., Sundareswaran, R., Just, J., Darr, M., & Jannesari, A. (2021). CLAWS: Contrastive learning with hard attention and weak supervision. *arXiv*. https://arxiv.org/abs/2112.00847

Jiang, Y., & Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: A review. *Plant Phenomics*, *2020*, 4152816. https://doi.org/10.34133/2020/4152816

Jing, L., & Tian, Y. (2021). Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(11), 4037–4058. https://doi.org/10.1109/TPAMI.2020.2992393

Jubery, T. Z., Carley, C. N., Singh, A., Sarkar, S., Ganapathysubramanian, B., & Singh, A. K. (2021). Using machine learning to develop a fully automated Soybean Nodule Acquisition Pipeline (SNAP). *Plant Phenomics*, *2021*, 9834746. https://doi.org/10.34133/2021/9834746

Kar, S., Nagasubramanian, K., Elango, D., Nair, A., Mueller, D. S., O'Neal, M. E., Singh, A. K., Sarkar, S., Ganapathysubramanian, B., & Singh, A. (2021). *Self-supervised learning improves agricultural pest classification*. Paper presented at the AI for Agriculture and Food Systems Workshop in the 2022 Association for the Advancement of Artificial Intelligence Conference. https://openreview.net/pdf?id=3ndIvP2491e

Khaki, S., Pham, H., Han, Y. e., Kuhl, A., Kent, W., & Wang, L. (2021). DeepCorn: A semi-supervised deep learning method for high-throughput image-based corn kernel counting and yield estimation. *Knowledge-Based Systems*, *218*(106874), 106874. https://doi.org/10.1016/j.knosys.2021.106874

Knight, P. A. (2008). The Sinkhorn–Knopp algorithm: Convergence and applications. *SIAM Journal on Matrix Analysis and Applications: A Publication of the Society for Industrial and Applied Mathematics*, *30*(1), 261–275. https://doi.org/10.1137/060659624

Kolesnikov, A., Zhai, X., & Beyer, L. (2019). Revisiting self-supervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1920–1929). Computer Vision Foundation.

Margapuri, V., & Neilsen, M. (2021). Classification of seeds using domain randomization on self-supervised learning frameworks. *arXiv*. https://arxiv.org/abs/2103.15578

Marin Zapata, P. A., Roth, S., Schmutzler, D., Wolf, T., Manesso, E., & Clevert, D.-A. (2021). Self-supervised feature extraction from image time series in plant phenotyping using triplet networks. *Bioinformatics*, *37*(6), 861–867. https://doi.org/10.1093/bioinformatics/btaa905

Marino, S., Beauseroy, P., & Smolarz, A. (2019). Weakly-supervised learning approach for potato defects segmentation. *Engineering Applications of Artificial Intelligence*, *85*, 337–346. https://doi.org/10.1016/j.engappai.2019.06.024

Misra, I., & van der Maaten, L. S. (2020). *Self-supervised learning of pretext-invariant representations*. Paper presented at the *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA. https://doi.org/10.1109/cvpr42600.2020.00674

Nagasubramanian, K., Jones, S., Sarkar, S., Singh, A. K., Singh, A., & Ganapathysubramanian, B. (2018). Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods*, *14*(1), 86. https://doi.org/10.1186/s13007-018-0349-9

Nagasubramanian, K., Jones, S., Singh, A. K., Sarkar, S., Singh, A., & Ganapathysubramanian, B. (2019). Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods*, *15*(1), 98. https://doi.org/10.1186/s13007-019-0479-8

Nagasubramanian, K., Jubery, T., Fotouhi Ardakani, F., Mirnezami, S. V., Singh, A. K., Singh, A., Sarkar, S., & Ganapathysubramanian, B. (2021). How useful is active learning for image-based plant phenotyping? *The Plant Phenome Journal*, *4*(1). https://doi.org/10.1002/ppj2.20020

Nagasubramanian, K., Singh, A. K., Singh, A., Sarkar, S., & Ganapathysubramanian, B. (2020). Usefulness of interpretability methods to explain deep learning based plant stress phenotyping. *arXiv*. https://arxiv.org/abs/2007.05729

Naik, H. S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., Singh, A., Singh, A. K., & Ganapathysubramanian, B. (2017). A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods*, *13*(1), 23. https://doi.org/10.1186/s13007-017-0173-7

Pantazis, O., Brostow, G., Jones, K., & Mac Aodha, O. (2021). Focus on the positives: Self-supervised learning for biodiversity monitoring. *arXiv*. https://arxiv.org/abs/2108.06435

Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2019). Machine learning approach for prescriptive plant breeding. *Scientific Reports*, *9*(1), 17132. https://doi.org/10.1038/s41598-019-53451-4

Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., & Singh, A. K. (2019). Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. *Plant Phenomics*, *2019*, 5809404.

Pérez-Ortiz, M., Peña, J. M., Gutiérrez, P. A., Torres-Sánchez, J., Hervás-Martínez, C., & López-Granados, F. (2015). A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Applied Soft Computing*, *37*, 533–544. https://doi.org/10.1016/j.asoc.2015.08.027

Pieruschka, R., & Schurr, U. (2019). Plant phenotyping: Past, present, and future. *Plant Phenomics*, *2019*, 7507131.

Pound, M. P., Atkinson, J. A., Townsend, A. J., Wilson, M. H., Griffiths, M., Jackson, A. S., Bulat, A., Tzimiropoulos, G., Wells, D. M., Murchie, E. H., Pridmore, T. P., & French, A. P. (2017). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*, *6*(10), 1–10. https://doi.org/10.1093/gigascience/gix083

Purushwalkam, S., Morgado, P., & Gupta, A. (2022). The Challenges of Continuous Self-Supervised Learning. *arXiv*. https://arxiv.org/abs/2203.12710

Rawat, S., Chandra, A. L., Desai, S. V., Balasubramanian, V. N., Ninomiya, S., & Guo, W. (2022). How useful is image-based active learning for plant organ segmentation? *Plant Phenomics*, *2022*, 9795275.

Ren, P., Xiao, Y., Chang, X., Huang, P. o- Y., Li, Z., Gupta, B. B., Chen, X., & Wang, X. (2022). A survey of deep active learning. *ACM Computing Surveys*, *54*(9), 1–40. https://doi.org/10.1145/3472291

Riera, L. G., Carroll, M. E., Zhang, Z., Shook, J. M., Ghosal, S., Gao, T., Singh, A., Bhattacharya, S., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2021). Deep multiview image fusion for soybean yield estimation in breeding applications. *Plant Phenomics*, *2021*, 9846470.

Settles, B. (2009). *Active learning literature survey*. https://minds.wisconsin.edu/handle/1793/60660

Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2021). Crop yield prediction integrating genotype and weather variables using deep learning. *Plos One*, *16*(6), e0252402. https://doi.org/10.1371/journal.pone.0252402

Shook, J. M., Zhang, J., Jones, S. E., Singh, A., Diers, B. W., & Singh, A. K. (2021). Meta-GWAS for quantitative trait loci identification in soybean. *G3*, *11*(7). https://doi.org/10.1093/g3journal/jkab117

Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science*, *21*(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015

Singh, A., Jones, S., Ganapathysubramanian, B., Sarkar, S., Mueller, D., Sandhu, K., & Nagasubramanian, K. (2021). Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends in Plant Science*, *26*(1), 53–69. https://doi.org/10.1016/j.tplants.2020.07.010

Singh, A. K., Singh, A., Sarkar, S., Ganapathysubramanian, B., Schapaugh, W., Miguez, F. E., Carley, C. N., Carroll, M. E., Chiozza, M. V., & Chiteri, K. O., & Others. (2021). High-throughput phenotyping in soybean. In J. Zhou & H. T. Nguyen (Eds.), *High-throughput crop phenotyping* (pp. 129–163). Springer.

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep learning for plant stress phenotyping: Trends and future perspectives. *Trends in Plant Science*, *23*(10), 883–898. https://doi.org/10.1016/j.tplants.2018.07.004

Singh, D. P., Singh, A. K., & Singh, A. (2021). *Plant breeding and cultivar development*. Academic Press.

Tian, Y., Sun, C., Poole, B., Krishnan, D., Schmid, C., & Isola, P. (2020). What makes for good views for contrastive learning? *Advances in Neural Information Processing Systems*, *33*, 6827–6839.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1). https://doi.org/10.1186/s40537-016-0043-6

Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. (2021). Barlow Twins: Self-supervised learning via redundancy reduction. *arXiv*. https://arxiv.org/abs/2103.03230

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE: Institute of Electrical and Electronics Engineers*, *109*(1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555

How to cite this article: Nagasubramanian, K., Singh, A., Singh, A., Sarkar, S., & Ganapathysubramanian, B. (2022). Plant phenotyping with limited annotation: Doing more with less. *The Plant Phenome Journal*, *5*, e20051. https://doi.org/10.1002/ppj2.20051