Computers and Electronics in Agriculture 175 (2020) 105542

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Few-Shot Learning approach for plant disease classification using images taken in the field

David Argüeso[a,b], Artzai Picon[a,*], Unai Irusta[b], Alfonso Medela[a], Miguel G San-Emeterio[a], Arantza Bereciartua[a], Aitor Alvarez-Gila[a]

[a] TECNALIA, Basque Research and Technology Alliance (BRTA), Parque Tecnológico de Bizkaia, C/ Geldo. Edificio 700, E-48160 Derio, Bizkaia, Spain
[b] Communications Engineering Department, University of the Basque Country, Alameda de Urquijo s/n, 48013 Bilbao, Bizkaia, Spain

ARTICLE INFO

Keywords:
Few-Shot Learning (FSL)
Plant disease
Fungal plant disease
Bacterial plant disease
Deep learning
Convolutional Neural Network (CNN)
Triplet loss
Contrastive loss

ABSTRACT

Prompt plant disease detection is critical to prevent plagues and to mitigate their effects on crops. The most accurate automatic algorithms for plant disease identification using plant field images are based on deep learning. These methods require the acquisition and annotation of large image datasets, which is frequently technically or economically unfeasible. This study introduces Few-Shot Learning (FSL) algorithms for plant leaf classification using deep learning with small datasets.

For the study 54,303 labeled images from the PlantVillage dataset were used, comprising 38 plant leaf and/or disease types (classes). The data was split into a source (32 classes) and a target (6 classes) domain. The Inception V3 network was fine-tuned in the source domain to learn general plant leaf characteristics. This knowledge was transferred to the target domain to learn new leaf types from few images. FSL using Siamese networks and Triplet loss was used and compared to classical fine-tuning transfer learning. The source and target domain sets were split into a training set (80%) to develop the methods and a test set (20%) to obtain the results. Algorithm performance was evaluated using the total accuracy, and the precision and recall per class. For the FSL experiments the algorithms were trained with different numbers of images per class and the experiments were repeated 20 times to statistically characterize the results.

The accuracy in the source domain was 91.4% (32 classes), with a median precision/recall per class of 93.8%/92.6%. The accuracy in the target domain was 94.0% (6 classes) learning from all the training data, and the median accuracy (90% confidence interval) learning from 1 image per class was 55.5 (46.0–61.7)%. Median accuracies of 80.0 (76.4–86.5)% and 90.0 (86.1–94.2)% were reached for 15 and 80 images per class, yielding a reduction of 89.1% (80 images/class) in the training dataset with only a 4-point loss in accuracy. The FSL method outperformed the classical fine tuning transfer learning which had accuracies of 18.0 (16.0–24.0)% and 72.0 (68.0–77.3)% for 1 and 80 images per class, respectively.

It is possible to learn new plant leaf and disease types with very small datasets using deep learning Siamese networks with Triplet loss, achieving almost a 90% reduction in training data needs and outperforming classical learning techniques for small training sets.

1. Introduction

Automatic plant disease classification algorithms are very important in agriculture, for example to prevent or mitigate plagues (Sladojevic et al., 2016; Al-Hiary et al., 2011; Rumpf et al., 2010; Sannakki et al., 2011). The rapid identification of plant diseases remains difficult in many parts of the world due to the lack of the necessary infrastructure. With current technologies images are available worldwide through high definition smart-phone cameras. So the development of low cost, general purpose, automatic, and accurate plant disease classification algorithms based on plant images taken in the field would make plant disease and plague control applications available worldwide.

Currently the best image classification results are obtained using deep learning algorithms based on convolutional neural networks (CNN), which have thousands or even millions of tuneable parameters. One of the limitations of these models is that they need large annotated datasets to learn the subtle visual features that characterize a disease in a plant image, or the subtle differences between similar plant species. The main drawback to develop automatic plant disease classification algorithms based on CNNs is the scarcity of data (images) for some

* Corresponding author.
E-mail address: artzai.picon@tecnalia.com (A. Picon).

https://doi.org/10.1016/j.compag.2020.105542
Received 31 January 2020; Received in revised form 21 April 2020; Accepted 29 May 2020
Available online 20 June 2020
0168-1699/ © 2020 Elsevier B.V. All rights reserved.

(a) Source domain, D_s (b) Target domain, D_t

Fig. 1. Examples from the 38 classes in the PlantVillage (Hughes and Salathé, 2015) dataset ordered (left–right) as in Table 1.

types of plant diseases. Some diseases are very infrequent in some crops, although their impact can be devastating causing plagues with very important economic consequences, specially in regions where those crops have been recently introduced (Martinelli et al., 2015). In addition, the process to accurately annotate a complete plant disease dataset is very time consuming, intensive in expert human labor, and thus expensive (Yang et al., 2017; Picon et al., 2019).

Images of leaves are one of the most important sources of information for the detection and classification of plant species and their diseases. Researchers have thus concentrated on developing automatic algorithms for the classification of leaf images (Mohanty et al., 2016). Several studies introduced classical computer vision techniques (Rumpf et al., 2010; Sannakki et al., 2011; Al-Hiary et al., 2011; Martinelli et al., 2015), others developed tools and applications like MedLeaf (Prasvita and Herdiyeni, 2013), a mobile application to identify medicinal plants. Meadleaf classifies 30 species of Indonesian medicinal plants with an accuracy of 56.33%. The model was developed using 48 sample images of each of the 30 species. Novotný and Suk (2013) proposed a system to recognize woody species in Central Europe using a dataset that included 151 species with at least 50 leaves per species. They achieved a success rate for 10 species of 99.63% using a kNN classifier. Other authors analyzed the detection of wheat diseases in the wild using classical techniques (Johannes et al., 2017), and embedded the algorithm in a mobile application. Recently, they have extended their algorithms to accurately identify fungal diseases on wheat at very early stages using CNNs (Picon et al., 2019).

There is an increasing trend to replace leaf image classification algorithms based on machine learning by algorithms based on deep learning methods (Mohanty et al., 2016; Sladojevic et al., 2016; Wang et al., 2017; Fuentes et al., 2017; Ferentinos, 2018; Geetharamani and Pandian, 2019; Chen et al., 2020; Zhong and Zhao, 2020). Unfortunately, the unavailability of large datasets with reliable annotations has hindered the development of such algorithms, specially for very specific plant species and/or diseases. Recent advances in deep learning solutions have proven the effectiveness of several architectures to learn new classes using small datasets, a subfield known as Few-Shot Learning (FSL) (Chen et al., 2019; Larochelle et al., 2020; Wang et al., 2019).

FSL methods for image classification include model initialization, metric learning and data hallucination methods. Initialization methods focus on learning good model initializations for the tunable parameters in the network, so that classifiers for new classes can be learned with a limited set of examples (Finn et al., 2018; Nichol and Schulman, 2018; Rusu et al., 2018). Metric learning methods learn to compare. The intuition is that once a network learns to compare classes, it will be able to learn new classes from few labeled instances (Koch et al., 2015). Several metric learning approaches have been explored including cosine similarity (Vinyals et al., 2016), euclidean distance to class-mean representations (Snell et al., 2017), CNN relation modules (Sung et al., 2018), ridge regression (Bertinetto et al., 2018), and graph neural networks (Kim et al., 2019). Finally, hallucination methods overcome data shortage by learning to augment the data. These methods learn a generator from the data in the base classes, and use that generator to hallucinate (generate) data for new classes. They can either transfer variance in a base class data to new classes (Hariharan and Girshick, 2017), or use generative adversarial networks (Zhang et al., 2018; Wang et al., 2020) or autoencoders (Schonfeld et al., 2019).

FSL solutions for plant leave classification have been introduced recently, using adversarial models (hallucination models) (Hu et al., 2019), and Siamese networks with Contrastive loss and kNN classifers (Wang and Wang, 2019). However, a FSL solution for plant leaf classification based on metric learning could be improved using more elaborate loss functions (Triplet loss) to learn features (Hermans et al., 2017), and better classifiers to more efficiently learn the decision boundaries from few examples (Medela et al., 2019). Consequently, the objective of this study was to demonstrate an FSL model for plant leaf classification based on Siamese networks and Triplet loss with an efficient class boundary learning algorithm based on multiclass support vector machines (SVM). For that purpose a general image classification algorithm was adapted for plant leaf classification using Contrastive and Triplet loss to learn from a large dataset. And then, the models were adapted to learn from few images of new plant leaf/disease classes, and

Table 1
Classes and number of images from the PlantVillage (Hughes and Salathé, 2015) dataset, grouped by source (D_s) and target (D_t) domains as used in this study, and k is the class number within the domain. An example of each class is shown in Fig. 1.

Crop	Disease	Images	D	k	Crop	Disease	Images	D	k
Cherry	powdery mildew	1052	D_s	0	Squash	powdery mildew	1835	D_s	19
Corn	cercospora	513	D_s	1	Strawberry	healthy	456	D_s	20
Corn	rust	1192	D_s	2	Strawberry	leaf scorch	1109	D_s	21
Corn	healthy	1162	D_s	3	Tomato	bacterial spot	2127	D_s	22
Corn	northern leaf blight	985	D_s	4	Tomato	early blight	1000	D_s	23
Grape	black rot	1180	D_s	5	Tomato	healthy	1592	D_s	24
Grape	black measles	1384	D_s	6	Tomato	late blight	1910	D_s	25
Grape	healthy	423	D_s	7	Tomato	leaf mold	952	D_s	26
Grape	isariopsis leaf spot	1076	D_s	8	Tomato	septoria leaf spot	1771	D_s	27
Orange	citrus greening	5507	D_s	9	Tomato	spider mites	1676	D_s	28
Peach	bacterial spot	2291	D_s	10	Tomato	target spot	1404	D_s	29
Peach	healthy	360	D_s	11	Tomato	mosaic virus	373	D_s	30
Pepper	bacterial spot	997	D_s	12	Tomato	Yellow leaf curl	5357	D_s	31
Pepper	healthy	1478	D_s	13	Apple	scab	630	D_t	0
Potato	early blight	1000	D_s	14	Apple	black rot	621	D_t	1
Potato	healthy	152	D_s	15	Apple	cedar rust	276	D_t	2
Potato	late blight	1000	D_s	16	Apple	healthy	1645	D_t	3
Raspberry	healthy	371	D_s	17	Blueberry	healthy	1502	D_t	4
Soybean	healthy	5090	D_s	18	Cherry	healthy	854	D_t	5

their performance was assessed and compared as function of the number of images used to learn the new classes.

2. Materials

In this study the PlantVillage project dataset was used (Hughes and Salathé, 2015). PlantVillage is a not-for-profit project led by Penn State University in the US and EPFL in Switzerland, in which tens of thousands of annotated plant leaf images of healthy and diseased plants have been annotated and made openly and freely available.

The PlantVillage dataset contains 54,303 leaf images with 38 classes that include 14 crop species and 26 diseases. All images have been collected under controlled conditions, both in grayscale and colour. For the experiments, all images were resized to 256 × 256 pixels, without any additional image preprocessing. Fig. 1 shows an example of the 38 classes, and Table 1 displays the types of crops and their diseases.

The classes were randomly divided to form two datasets: a source domain D_s with 32 classes (48,775 images) to develop the baseline plant leaf classification algorithm, and a target domain D_t with the remaining 6 classes (5528 images) to develop and evaluate the FSL algorithms. Both datasets were split into a training set (80%) and a test set (20%). All the data partitions were made randomly and in a stratified way (preserving the train/test proportions on each class). When more than one image came from the same leaf but at different orientations and/or conditions, all those images were forced into the same partition to avoid data leakage.

3. Methods

The architecture of the plant leaf image classification algorithm is based on a recent proposal by Medela et al. (2019) and Medela and Picon (2019), and is shown in Fig. 2. It consists of two blocks. First, a general purpose CNN image classification network was fine tuned to extract leaf image features or image embeddings. The CNN learns how to map a plant leaf image X_i to $f_i = f(X_i)$, a low dimensional feature vector or mapping. Second, a shallow SVM classifier was trained to learn the distances between the f_i mappings for the different plant leaf classes, and thus learn to separate the classes. This algorithm was trained with the source domain dataset, D_s.

The hypothesis was that the image mappings and distance representations learned by the algorithm in dataset D_s would be general enough, so that knowledge could be transferred to classify new plant leaf classes using few images of those new classes. This idea is

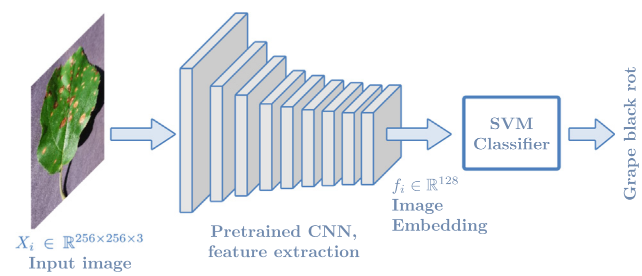

Fig. 2. Block diagram of the plant leaf classification algorithm. The algorithm has two blocks. First, a general purpose image feature extraction CNN adapted to plant leaf images that embeds an image X_i as a feature vector $f_i = f(X_i)$. Second, a SVM classifier trained to separate classes based on the f_i image embeddings.

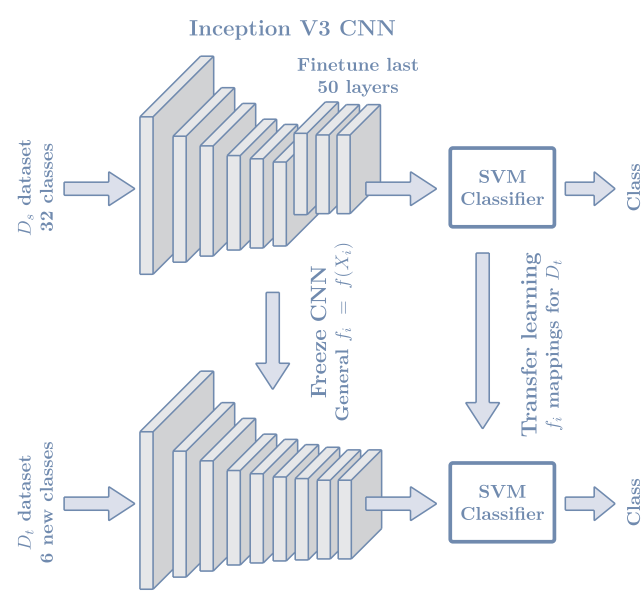

Fig. 3. Schematics of the FSL architectures. In the upper branch a general plant leaf feature extraction CNN is trained by fine-tuning the Inception V3 network. Then, the trained feature extractor is used to create class-distance embeddings of the images of a new small dataset, and these feature maps are used to fine tune a shallow SVM classifier.

schematically shown in Fig. 3, which shows the FSL architecture. The upper branch is the general purpose plant leaf classifier trained on dataset D_s. The lower branch is the FSL algorithm which is formed by freezing the CNN image mapping network, and by readjusting the class distance representation of the SVM classifier. This knowledge transfer is done using a smaller dataset, in the experiments the dataset was a subset of D_t. The challenge is then to determine the number of images per class needed from D_t for an efficient transfer of the knowledge gained from D_s.

3.1. Baseline CNN

For the baseline CNN the Inception V3 network pretrained with the ImageNet database was used (Deng et al., 2009). The Inception V3 architecture is a refinement of the original Inception V1 (Szegedy et al., 2015) that includes batch normalization and factorization (Szegedy et al., 2016). The network contains 315 layers, and to fine tune the network for plant leaf classification the first 265 layers were frozen, and the last 50 layers were trained with the D_s dataset. Once the network had learned to extract the image embeddings f_i, the fully connected layers were removed and replaced by the SVM classifier. After some preliminary experiments an embedding dimension of $N = 128$ was selected, so the CNN network assigns the image embedding $f_i \in \mathbb{R}^{128}$ to a color image $X_i \in \mathbb{R}^{256\times256\times3}$.

3.2. Siamese learning architectures

Two training architectures were used to learn the image embeddings: (1) a Siamese network with two subnets and Contrastive loss, and (2) a Siamese network with three subnets and Triplet loss. In Siamese networks all the subnets in the architecture share the same weights (see Fig. 4). For comparative purposes, the subnets of the Siamese networks were the same as in the baseline fine-tuning model, i.e. the Inception V3 networks with 50 tunable layers.

In the two subnet Siamese network a pair of images X_i and X_j is fed to the network during training, and the Contrastive loss function for the pair is calculated as:

$$\mathscr{L}_c(X_i, X_j) = y\cdot\|f_i - f_j\|^2 + (1-y)\cdot\max(0,\ m - \|f_i - f_j\|^2) \tag{1}$$

where $\|\cdot\|^2$ represents the euclidean distance, m is the margin, and $y = 0$ if X_i and X_j are of different class and $y = 1$ if they are of the same class. By minimizing the loss function, the network will learn to minimize the distance between embeddings for similar classes ($y = 1$), and maximize the distance between different classes ($y = 0$) up to the margin m.

In the three subnet Siamese network three images are fed to the network during training, an anchor, a positive and a negative image, that is X_a, X_p and X_n. The anchor and positive images are of the same class, and the negative image of a different class. The Triplet loss function is then:

$$\mathscr{L}_t(X_a, X_p, X_n) = \max(0,\ m + \|f_a - f_p\|^2 - \|f_a - f_n\|^2). \tag{2}$$

By minimizing the loss function the distance between the embeddings for the anchor and positive images is minimized, while maximizing the distance between the embeddings for the anchor and the negative class up to the margin m.

3.3. SVM classifier

The SVM classifier is designed to maximize the margin (gap) between the feature map representations of different classes (Cortes and Vapnik, 1995). SVM classifiers were originally conceived for binary classification problems, but can be generalized to multiclass classification in different ways (Hsu and Lin, 2002). For the proposed architecture, a one versus all multiclass SVM classifier was used. This means that for a K-class problem K different SVM classifiers are trained and a winner takes all strategy is followed to decide the class. Each classifier takes one class as positive and the rest of the classes as negative, and the class for embedding f_i is decided as:

$$c(f_i) = \underset{k}{\operatorname{argmax}}\{d_k(f_i)\} \tag{3}$$

where $d_k(\cdot)$ is the decision function for the SVM designed for class $k \in \{1, \ldots K\}$. For the experiments SVMs with linear kernels were used and the soft margin of the SVM was set to $C = 1$ (Medela et al., 2019; Pedregosa et al., 2011).

3.4. Few-shot and baseline transfer learning

Finally, to transfer the knowledge from the source domain D_s to the target domain D_t, the SVM classifier was retrained with the image embeddings of the target domain. The image embeddings for each of the three methods were obtained from the CNN networks trained on D_s. At this stage, only the shallow classifier was retrained with the training data from D_t. Different amount of training images per class were considered in the experiments, from $L = 1$ to $L = 140$ images per class. A typical fine-tuning transfer learning was used to evaluate the effectiveness of the FSL approach. There, the last 50 layers of the Inception V3 network were first trained on the D_s dataset, and then fine-tuned using the subsets of the D_t dataset. This baseline solution was named fine-tuning, and the FSL approaches using Siamese networks were named Contrastive loss and Triplet loss, respectively. In all these experiments the test set in D_t was fixed to 50 images per class, and the process was repeated 20 times to statistically characterize the accuracy of the methods. The accuracies of the different methods were compared using the McNemar test at the 95% confidence interval, so a value of

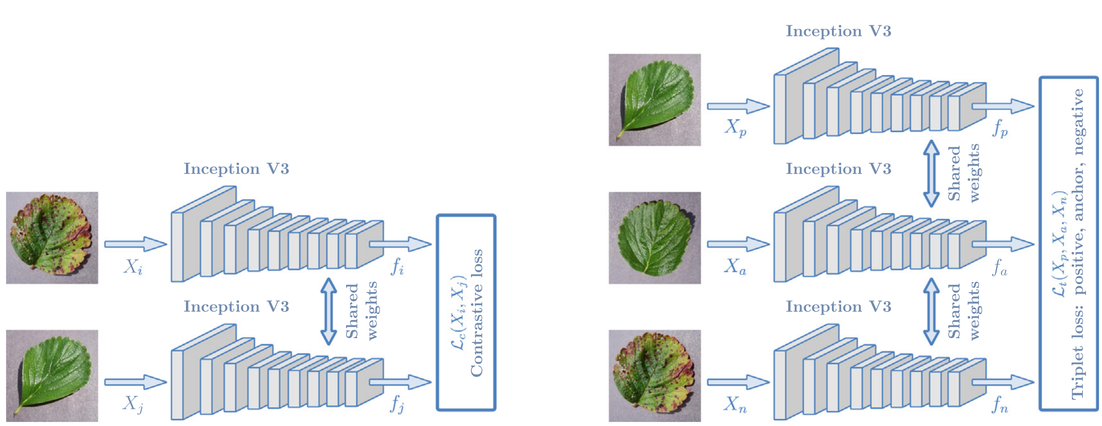

Fig. 4. Architecture of the Siamese networks. On the right the two-headed Siamese network based on Contrastive loss ($\mathscr{L}_c$), on the left the three-headed Siamese network based on Triplet loss ($\mathscr{L}_t$).

$p < 0.05$ was considered significant (Fagerland et al., 2013).

3.5. Network optimization

In all the experiments the training sets were augmented by rotating the images by a randomly selected angle from the set {0, 90, 180, 270}. A batch size of 32 was used to train the networks, with 30 epochs and a learning rate of $\alpha = 10^{-4}$. All these values were determined in some preliminary experiments. For the fine-tuning solution an Adam optimizer and the categorical cross-entropy loss function were used. For the Contrastive loss solution the Root Mean Square optimizer (RMSprop) was used, with a margin of $m = 1$ in the loss function. For the Triplet loss solution an Adam optimizer was used, with a margin of $m = 1$ in the loss function, and the pixel intensities were normalized to the $[-1, 1]$ range. All experiments were conducted on the Keras framework (Chollet et al., 2015) with a Tensorflow backend (Abadi et al., 2015).

3.6. Evaluation

Plant leaf image classification is a multiclass classification problem, so let us call N_{ij} the number of images of class i (true label) classified by the algorithm as class j (predicted label). In this case $i, j \in \{1, ...32\}$ for the main branch in the D_s dataset, and $i, j \in \{1, ...6\}$ for the FSL problem in the D_t dataset. The N_{ij} numbers can be arranged in a matrix (confusion matrix) from which the recall (R_i), precision (P_i) and F-score ($F_{1,i}$) for class i can be computed as:

$$\mathrm{R}_i = \frac{N_{ii}}{\sum_j N_{ij}}, \quad \mathrm{P}_i = \frac{N_{ii}}{\sum_i N_{ij}}, \quad F_{1,i} = 2\frac{\mathrm{P}_i \cdot \mathrm{R}_i}{\mathrm{P}_i + \mathrm{R}_i} \tag{4}$$

Finally, the multiclass accuracy (Acc), or its converse the error rate (err), can be calculated as summarizing metrics for all the classes:

$$\mathrm{Acc} = \frac{\sum_i N_{ii}}{\sum_{i,j} N_{ij}}, \quad \mathrm{err} = 1 - \mathrm{Acc} \tag{5}$$

In the experiments in the D_t dataset classes were well balanced, so the multiway accuracy is a good overall measure of accuracy (Rad et al., 2017).

All the metrics are reported as median and 90% confidence interval (CI) for the 20 random repetitions of the experiments, and a stacked (summed) confusion matrix is presented as summary.

4. Results

The 32-class models in the test subset of the source domain had accuracies of 91.4% and 90.6% for the Triplet and the Contrastive loss Siamese networks, respectively. Moreover, the median (90% CI) precision, recall and F-score per class were 93.8 (77.4–99.3)%, 92.6 (65.8–99.4)% and 92.3 (69.9–98.8)% for the Triplet loss; and 93.8 (73.8–99.1)%, 93.1 (54.1–99.9)% and 91.2 (64.8–98.8)% for the Contrastive loss. Not only did the Siamese architectures learn how to classify leaf types and disease types, but they did it for a vast majority of the 32 classes. The number of classes with F-scores below 70% was 3 for Triplet loss and 5 for Contrastive loss. And as shown in Fig. 5 all classes had recall, precision and F-score values above 55.0% for the Triplet loss, which is a considerable improvement over the 3.1% accuracy for a random guess in 32 classes.

The median (90% CI) accuracy for the FSL approaches as a function of the images per class (L) in the D_t dataset is shown in Fig. 6. All curves increase sharply as L grows until the methods reach a plateau in accuracy. However, the accuracy of the FSL methods was significantly higher ($p > 0.05$) than that of the fine tuning model, except for Contrastive loss with $L > 100$. For $L > 5$ the FSL method based on Triplet loss had significantly higher accuracies ($p < 0.05$) than that based on Contrastive loss.

For the Triplet loss method 80% and 90% accuracies were reached for as few as $L = 15$ and $L = 80$ (see Fig. 6). In fact, when all 4421 training samples from D_t were used the accuracy was 94.0%, only 4-points larger than the 90.0 (86.1–94.2)% obtained with Triplet loss and $L = 80$ (480 images for the 6 classes). This means a reduction of 89.1% in the data set size requirements with a loss of only 4-points in accuracy. For the Triplet loss, the accuracy (or conversely the error) converges rapidly to the one obtained using the complete training D_t dataset, as shown in Fig. 7. The results for Contrastive loss were worse, reaching an 80% accuracy for $L = 80$. In fact, the accuracy for Contrastive loss with the complete D_t training dataset was 89.6%, below that of Triplet loss with $L = 80$. And not only this upper limit was lower, or the error higher at 10.4%, but the method converged to that limit more slowly than for the Triplet loss (see Fig. 7).

An interesting insight into how the best method performed on the

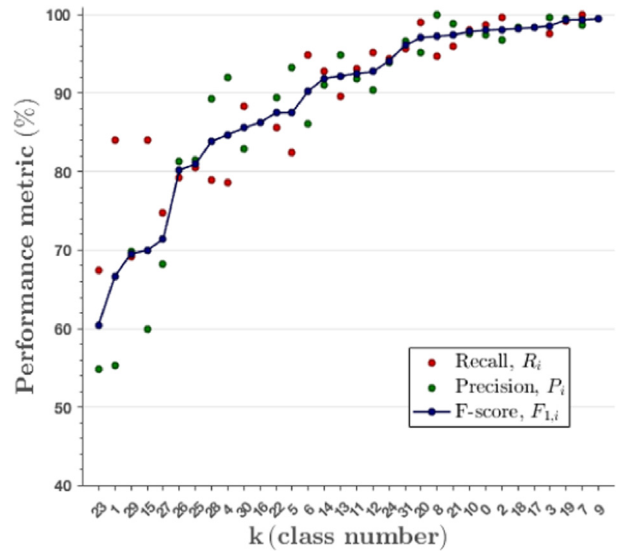

Fig. 5. Performance metrics per class for the Triplet loss model in the test set of the D_s domain. The classes (horizontal axis, see Table 1) are sorted by ascending F-score values.

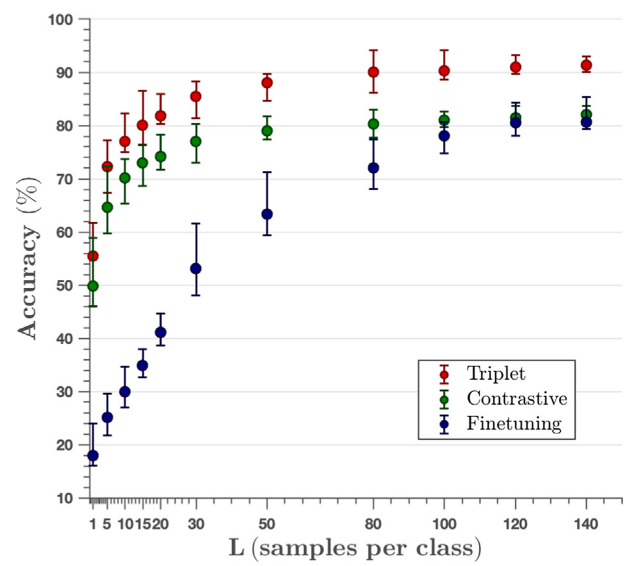

Fig. 6. Accuracy in D_t (6 classes) as a function of the number of training samples per class. The graphs show the median (90% CI) for the experiments with the FSL approaches, Triplet (red) and Contrastive loss (green), compared to the baseline fine-tuning model (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

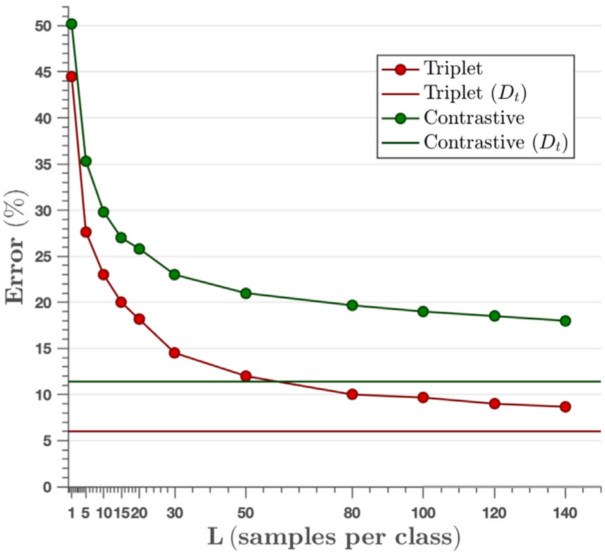

Fig. 7. Median error rate as a function of the samples per class L, for the FSL methods. The error rates are compared to the theoretical limits attainable by each method when the complete training D_t was used, that is 4421 images or on average $L = 737$ images per class.

target domain is to analyze the stacked confusion matrices for D_t with Triplet loss for low and moderate values of L. This is shown in Fig. 8. When few samples were used to train the model ($L = 10$) the overall accuracy was 77.0 (75.0–82.3)%. The network had difficulties differentiating leaves with very similar morphologies such as (see Fig. 9): apple scab (class 0), apple black rot (1), cedar apple rust (2) and healthy apple leaves (3). All four leaves correspond to the same crop in different varieties and/or disease conditions. When compared to leaves of different morphologies, like healthy blueberry (4) or cherry leaves (5), errors were very infrequent, the F-scores for those two classes were 86.9% and 90.0%, respectively. However, when the samples used to train the model were increased to $L = 80$ the lowest F-score was 79.4% for apple scab, and the rest of the classes had F-scores above 85%, and 4 classes ($\{1, 2, 4, 5\}$) had F-scores above 90%

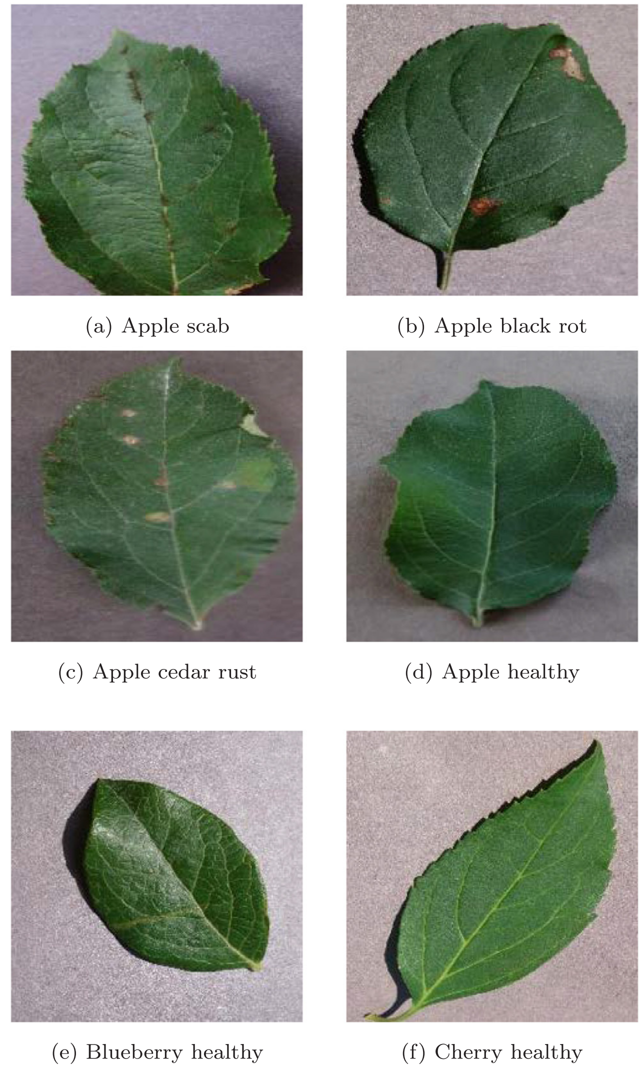

Fig. 9. Examples of leaves in the target domain, with similar morphologies (top four) or different morphologies (lower two).

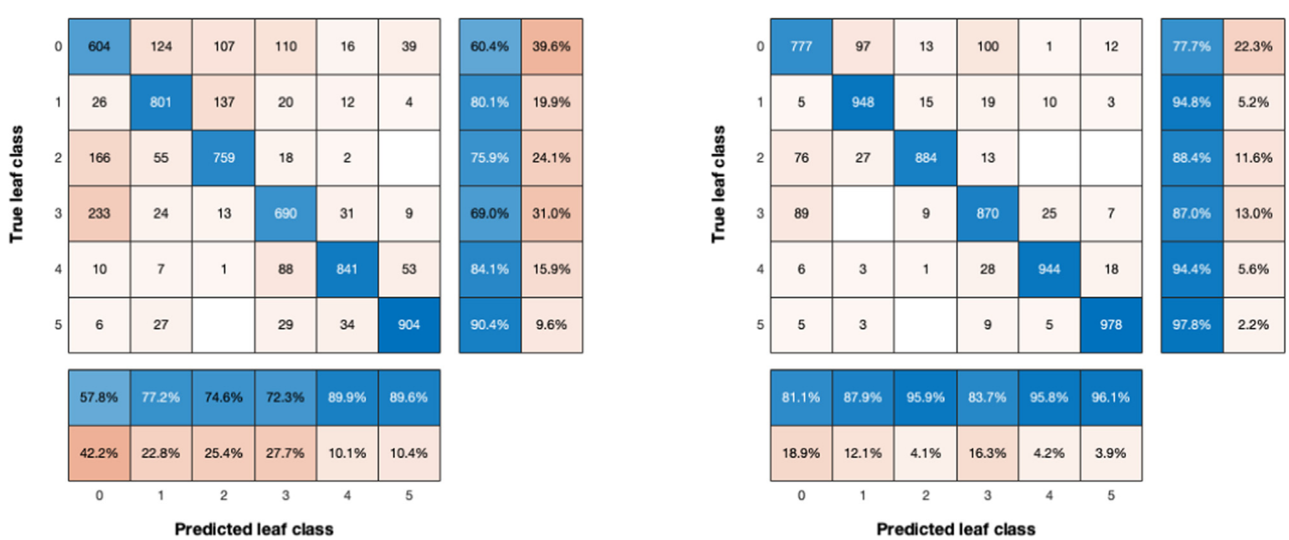

Fig. 8. Stacked confusion matrices for the 20 random experiments with the Triplet loss, for $L = 10$ samples per class (left) and $L = 80$ (right). The bottom rows show the recall (R_i) per class and the leftmost columns the precision (P_i) per class. The test set contained 50 samples per class or a total of 6000 decisions in the 20 runs. For the class labels in D_t consult Table 1 or Fig. 9.

5. Discussion

Intensive dataset annotation and manual labeling have been common practice on deep-learning based applications in agriculture (Kamilaris and Prenafeta-Boldu, 2018). In plant disease classification tasks the PlantVillage dataset is a good example, with 54,303 images taken under laboratory conditions (Mohanty et al., 2016). Over real field conditions, Johannes et al. (2017) used 3637 images to detect three distinct wheat diseases, and had to extend their dataset to 8,178 images to allow the detection of the early stages of the diseases. Moreover, broadening the scope of those methods to allow 5 crops and 17 diseases involved the acquisition and annotation of 121,955 images (Picon et al., 2019), a process that spanned a three-year campaign. Scaling the datasets to include new diseases is thus time-consuming and suboptimal. The problem is aggravated in the case of general purpose plant leaf classifiers with user fed datasets, where the number of classes keeps increasing and the quality and quantity of the images are variable depending on the prevalence of the diseases and the user's interests. In such scenarios usual deep learning approaches are unpractical. But, as our results demonstrate, a distance metric approach allows learning from few examples, and is therefore a realizable solution cost and time-wise, with just a slight decrease in accuracy.

Our results show that adapting a general purpose image classification network like Inception V3 for plant leaf classification is a viable solution. The baseline accuracies for the 32-class problem in the source domain were 90.64% and 91.38% for the Siamese networks with Contrastive and Triplet loss, respectively. Moreover, all leaf types and diseases were identified with precisions above 50%, and F-scores above 60% (see Fig. 5). However, the networks trained in the source domain still required a large dataset of annotated images, around 1200 images per class in the training set of D_s. The key question was then to see how well these results generalized to the target domain with few images. When trained with a single image per class in D_t the baseline transfer method had an accuracy of 18.0 (16.0–24.0)%, which is similar to 16.7%, the accuracy of a random guess for a 6-class problem. However, with Contrastive or Triplet loss the accuracies were at 49.8 (46.0–58.9)% and 55.5 (46.0–61.7)%, respectively, which shows that knowledge was transferred very efficiently to the target domain with as few as a single example per class. Moreover, using the Triplet loss and 80 images per class the accuracy was 90.0 (86.1–94.2)%, only 4 points below the accuracy obtained using the complete training dataset. There is a trade-off between accuracy and the cost/lead-times to develop accurate algorithms for new plant leaf and disease classifications. The approach introduced in this study produces very accurate algorithms with few image examples, i.e. usable algorithms with very short lead times. Moreover, those algorithms would improve their performance as more labeled data becomes available.

The FSL architecture designed for this study is based on the Inception V3 network and a linear SVM classifier. Other general purpose embedding extraction networks could also be used such as VGGNet (Simonyan and Zisserman, 2014; Koch et al., 2015; Medela and Picon, 2019) or Resnet (He et al., 2016). Some of these networks have already been used in the main applications of FSL based on Siamese networks, such as face recognition (Kumari et al., 2019) and person re-identification (Hermans et al., 2017; Lv et al., 2019), speaker verification (Bredin, 2017) or object retrieval (He et al., 2018). In fact, Wang and Wang (2019) compared the Inception, ResNet and DenseNet architectures for a plant leaf FSL solution based on contrastive loss, and showed that the Inception network produced the highest classification accuracies for three other publicly available leaf classification datasets. Our results show that the FSL solution can be further improved introducing the Triplet loss to learn the image embeddings, and a more elaborate multiclass SVM to efficiently learn the new class boundaries from few examples. Siamese networks based on Triplet loss overcome the convergence problems Siamese networks with Contrastive loss because by using a positive and a negative image at each iteration Triplet loss avoids local minima during training (Medela and Picon, 2019). The one versus all SVM model proved efficient at learning new class boundaries from few examples. However, as the number of classes and the class imbalances increase, other more elaborate classification approaches may be needed (Wang and Yao, 2012; Zhang et al., 2016). Such situations may arise in general plant species classification, or rare plant disease identification.

6. Conclusions

This study shows that a distance metric FSL approach based on the Triplet loss and an efficient class boundary shallow learner improves current methods for FSL plant leaf classification. After training a general purpose CNN to learn to extract general plant leaf characteristics, our approached showed accuracy above 90% using as few as 80 images per class on six new species/diseases. The training dataset size could be reduced by 89.1% with just a 4-point loss in accuracy. This demonstrates that it is possible to develop accurate new algorithms to identify new plant species and diseases with very few annotated training images. These FSL methods would substantially shorten the production time and reduce the costs of developing new plant disease identification algorithms.

CRediT authorship contribution statement

David Argüeso: Conceptualization, Formal analysis, Investigation, Software, Writing - original draft. **Artzai Picon:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing - original draft, Writing - review & editing. **Unai Irusta:** Conceptualization, Investigation, Writing - original draft, Writing - review & editing. **Alfonso Medela:** Conceptualization, Formal analysis, Investigation, Writing - review & editing. **Miguel G San-Emeterio:** Conceptualization, Validation, Writing - review & editing. **Arantza Bereciartua:** Conceptualization, Investigation, Methodology, Writing - review & editing. **Aitor Alvarez-Gila:** Conceptualization, Investigation, Methodology, Writing - review & editing.

Acknowledgements

This research was funded by the ELKARTEK Research Programme of the Basque Government. Project #KK-2019/00068 and through grant IT-1229-19.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.compag.2020.105542.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M., ALRahamneh, Z., 2011. Fast and accurate detection and classification of plant diseases. International Journal of Computer Applications 17 (1), 31–38.

Bertinetto, L., Henriques, J.F., Torr, P.H., Vedaldi, A., 2018. Meta-learning with differentiable closed-form solvers, arXiv preprint arXiv:1805.08136.

Bredin, H., 2017. Tristounet: triplet loss for speaker turn embedding. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 5430–5434.

Chen, W., Liu, Y., Kira, Z., Wang, Y.F., Huang, J., 2019. A closer look at few-shot classification. CoRR, vol abs/1904.04232.

Chen, J., Chen, J., Zhang, D., Sun, Y., Nanehkaran, Y., 2020. Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture 173, 105393.

Chollet, F. et al., 2015. Keras, URL:https://keras.io.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20 (3), 273–297.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255.
Fagerland, M.W., Lydersen, S., Laake, P., 2013. The mcnemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. BMC Medical Research Methodology 13, 91.
Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture 145, 311–318.
Finn, C., Xu, K., Levine, S., 2018. Probabilistic model-agnostic meta-learning. In: Advances in Neural Information Processing Systems, pp. 9516–9527.
Fuentes, A., Yoon, S., Kim, S., Park, D., 2017. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors 17 (9), 2022.
Geetharamani, G., Pandian, A., 2019. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Computers & Electrical Engineering 76, 323–338.
Hariharan, B., Girshick, R., 2017. Low-shot visual recognition by shrinking and hallucinating features. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3018–3027.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 770–778.
He, X., Zhou, Y., Zhou, Z., Bai, S., Bai, X., 2018. Triplet-center loss for multi-view 3d object retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1945–1954.
Hermans, A., Beyer, L., Leibe, B., 2017. In defense of the triplet loss for person re-identification, arXiv preprint arXiv:1703.07737.
Hsu, C.-W., Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. IEEE Transactions on Neural Networks 13 (2), 415–425.
Hu, G., Wu, H., Zhang, Y., Wan, M., 2019. A low shot learning method for tea leaf's disease identification. Computers and Electronics in Agriculture 163, 104852.
Hughes, D.P., Salathé, M., 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. CoRR, vol abs/1511.08060.
Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A.D., Ortiz-Barredo, A., 2017. Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. Computers and Electronics in Agriculture 138, 200–209.
Kamilaris, A., Prenafeta-Boldu, F.X., 2018. Deep learning in agriculture: a survey. Computers and Electronics in Agriculture 147, 70–90.
Kim, J., Kim, T., Kim, S., Yoo, C.D., 2019. Edge-labeling graph neural network for few-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11–20.
Koch, G., Zemel, R., Salakhutdinov, R., 2015. Siamese neural networks for one-shot image recognition. In: ICML Deep Learning Workshop, vol. 2.
Kumari, S., Pandey, P., Deshmukh, M., 2019. Face tagging and recognition using inception network and triplet loss generator function. In: International Conference on Communication and Intelligent Systems. Springer, pp. 161–174.
Larochelle, H., 2020.. Few-Shot Learning, ch. Computer Vision: A Reference Guide, Springer International Publishing, pp. 1–4.
Lv, X., Zhao, C., Chen, W.., 2019. A novel hard mining center-triplet loss for person re-identification. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Springer, pp. 199–210.
Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L.R., et al., 2015. Advanced methods of plant disease detection. a review. Agronomy for Sustainable Development 35 (1), 1–25.
Medela, A., Picon, A., 2019. Constellation loss: Improving the efficiency of deep metric learning loss functions for optimal embedding, arXiv preprint arXiv:1905.10675.
Medela, A., Picon, A., Saratxaga, C.L., Belar, O., Cabezon, V., Cicchi, R., Bilbao, R., Ben, G., 2019. Few shot learning in histopathological images: reducing the need of labeled data on biological datasets. In: IEEE International Symposium on Biomedical Imaging.
Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Frontiers in Plant Science 7, 1419.
Nichol, A., Schulman, J., 2018. Reptile: a scalable metalearning algorithm, , vol. 2, arXiv preprint arXiv:1803.02999.
Novotný, P., Suk, T., 2013. Leaf recognition of woody species in central Europe. Biosystems Engineering 115 (4), 444–452.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in python. Journal of Machine Learning Research 12, 2825–2830.
Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., Johannes, A., 2019. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Computers and Electronics in Agriculture 161, 280–290.
Picon, A., Seitz, M., Alvarez-Gila, A., Mohnke, P., Ortiz-Barredo, A., Echazarra, J., 2019. Crop conditional convolutional neural networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. Computers and Electronics in Agriculture 167, 105093.
Prasvita, D.S., Herdiyeni, Y., 2013. Medleaf: mobile application for medicinal plant identification based on leaf image. International Journal on Advanced Science, Engineering and Information Technology 3 (2), 103–106.
Rad, A.B., Eftestøl, T., Engan, K., Irusta, U., Kvaløy, J.T., Kramer-Johansen, J., Wik, L., Katsaggelos, A.K., 2017. ECG-based classification of resuscitation cardiac rhythms for retrospective data analysis. IEEE Transactions on Biomedical Engineering 64 (10), 2411–2418.
Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., Plümer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and Electronics in Agriculture 74 (1), 91–99.
Rusu, A.A., Rao, D., Sygnowski, J., Vinyals, O., Pascanu, R., Osindero, S., Hadsell, R., 2018. Meta-learning with latent embedding optimization, arXiv preprint arXiv:1807.05960.
Sannakki, S.S., Rajpurohit, V.S., Nargund, V., Kumar, A., Yallur, P.S., 2011. Leaf disease grading by machine vision and fuzzy logic. International Journal of Computer Technology and Applications 2 (5), 1709–1716.
Schonfeld, E., Ebrahimi, S., Sinha, S., Darrell, T., Akata, Z., 2019. Generalized zero-and few-shot learning via aligned variational autoencoders. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8247–8255.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.
Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D., 2016. Deep neural networks based recognition of plant diseases by leaf image classification. Computational Intelligence and Neuroscience 2016.
Snell, J., Swersky, K., Zemel, R., 2017. Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems, pp. 4077–4087.
Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M., 2018. Learning to compare: relation network for few-shot learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1199–1208.
Szegedy, C., Liu, Wei, Jia, Yangqing, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1–9.
Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826.
Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al., 2016. Matching networks for one shot learning. In: Advances in Neural Information Processing Systems, pp. 3630–3638.
Wang, B., Wang, D., 2019. Plant leaves classification: a few-shot learning method based on siamese network. IEEE Access 7, 151754–151763.
Wang, S., Yao, X., 2012. Multiclass imbalance problems: analysis and potential solutions. IEEE Transactions on Systems, Man, and Cybernetics Part B (Cybernetics) 42 (4), 1119–1130.
Wang, G., Sun, Y., Wang, J., 2017. Automatic image-based plant disease severity estimation using deep learning. Computational Intelligence and Neuroscience 2017.
Wang, Y., Yao, Q., Kwok, J., Ni, L.M., 2019. Generalizing from a few examples: a survey on few-shot learning, arXiv:1904.05046.
Wang, T., Ding, Z., Shao, W., Tang, H., Huang, K., 2020. Towards fair cross-domain adaptation via generative learning, arXiv preprint arXiv:2003.02366.
Yang, L., Zhang, Y., Chen, J., Zhang, S., Chen, D.Z., 2017. Suggestive annotation: a deep active learning framework for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 399–407.
Zhang, Z., Krawczyk, B., Garcìa, S., Rosales-Pérez, A., Herrera, F., 2016. Empowering one-vs-one decomposition with ensemble learning for multi-class imbalanced data. Knowledge-Based Systems 106, 251–263.
Zhang, R., Che, T., Ghahramani, Z., Bengio, Y., Song, Y., 2018. Metagan: an adversarial approach to few-shot learning. In: Advances in Neural Information Processing Systems, pp. 2365–2374.
Zhong, Y., Zhao, M., 2020. Research on deep learning in apple leaf disease recognition. Computers and Electronics in Agriculture 168, 105146.