Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 13(3), 1482-1495, 2023
Journal of the Institute of Science and Technology, 13(3), 1482-1495, 2023

ISSN: 2146-0574, eISSN: 2536-4618

Computer Engineering **DOI: 10.21597/jist.1283491**

Research Article

Received: 14.04.2023 **Accepted: 05.05.2023**

To Cite: Işık, G. (2023). Improving Plant Disease Recognition Through Gradient-Based Few-shot Learning with Attention Mechanisms. *Journal of the Institute of Science and Technology*, 13(3), 1482-1495.

Improving Plant Disease Recognition Through Gradient-Based Few-shot Learning with Attention Mechanisms

Gültekin IŞIK*

Highlights:

- Gradient based meta-learning algorithm used for few-shot classification
- Attention modules used for feature extraction
- Classifiers used in combination through a soft voting technique

Keywords:

- Few-shot learning
- Meta-learning
- Attention mechanisms
- Plant diseases
- Deep learning

ABSTRACT:

This study investigates the use of few-shot learning algorithms to improve classification performance in situations where traditional deep learning methods fail due to a lack of training data. Specifically, we propose a few-shot learning approach using the Almost No Inner Loop (ANIL) algorithm and attention modules to classify tomato diseases in the Plant Village dataset. The attended features obtained from the five separate attention modules are classified using a Multi Layer Perceptron (MLP) classifier, and the soft voting method is used to weigh the classification scores from each classifier. The results demonstrate that our proposed approach achieves state-of-the-art accuracy rates of 97.05% and 97.66% for 10-shot and 20-shot classification, respectively. Our approach demonstrates the potential for incorporating attention mechanisms in feature extraction processes and suggests new avenues for research in few-shot learning methods.

*Gültekin IŞIK, (Orcid ID: 0000-0003-3037-5586), Iğdır University, Department of Computer Engineering, Iğdır, Türkiye

Corresponding Author: Gültekin IŞIK, e-mail: gultekin.isik@igdir.edu.tr

INTRODUCTION

Plant diseases pose a seriously threatening to global food security, with significant impacts on crop yield and quality. Early detection and accurate diagnosis of plant diseases are crucial for effective disease management and prevention of crop losses. Relying on the manual monitoring of plant diseases is often associated with inaccuracies, and hiring domain experts for this purpose can be a challenge for farmers due to the high costs involved. Therefore, the development of an intelligent plant disease diagnosis system has become a necessity in the agricultural sector for the regular monitoring of crop fields. In this context, the automatic classification of plant leaf diseases has emerged as an important area of research (Albattah et al., 2022). The identification of plant diseases through leaf images has become a key focus of research in precision agriculture, where automatic classification algorithms are essential to prevent or mitigate the impact of pest infestations (Patricio & Rieder, 2018). However, the data scarcity problem is a major challenge in plant disease recognition, as it is difficult to collect a large and diverse set of labeled data for each plant disease. This limits the effectiveness of traditional machine learning approaches, which rely on large amounts of labeled data for training that can be time-consuming and expensive to acquire (S. Wang et al., 2021). Few-shot learning is a subfield of machine learning that addresses this challenge by enabling models to learn from a few labeled examples. As reported by (Yang et al., 2022), agriculture ranks third in terms of the prevalence of few-shot learning applications, following the medical and biological domains.

Few-shot learning can be defined as the process of learning from a limited amount of labeled data, typically ranging from 1 to 10 examples per class. The challenge in few-shot learning is to generalize well to new, unseen examples, given the limited amount of labeled data. Several approaches have been proposed to address this problem, including transfer learning (Sun et al., 2019), self-supervised learning (Liu et al., 2021), and meta-learning (Finn et al., 2017; Nichol et al., 2018; Raghu et al., 2019). In this study, we focus on the meta-learning approach, which has shown promising results in few-shot learning.

Meta-learning, also known as learning-to-learn, is a popular approach for few-shot learning that aims to learn a meta-model that can adapt to new tasks with only a few examples. The idea is to train a model on a variety of related tasks, such that it can learn to quickly adapt to new tasks with limited data. In the context of few-shot classification, the goal is to learn a model that can classify new examples with only a few labeled examples per class. In the meta-learning paradigm, it is crucial to have a significant number of tasks for adaptation. The model-based approach relies on an adaptive internal state rather than a fixed neural network when testing the model. This approach entails the utilization of a stateful and internal representation of a task that captures pertinent task-specific information, which is then utilized to make predictions for new inputs (Munkhdalai & Yu, 2017). Gradient-based methods aim to learn a set of initial parameters that can quickly adapt to new tasks. This involves optimizing the model's parameters with respect to the loss on the training set of a particular task, and the learned parameters are then utilized to perform well on unseen tasks (Finn et al., 2017; Nichol et al., 2018; Raghu et al., 2019). Metric-based meta-learning is another approach that involves learning a metric to compare instances of different classes, which is then used to make predictions (Snell et al., 2017). These approaches have been widely used in few-shot learning tasks, including agricultural tasks such as plant disease classification.

Several gradient-based meta-learning algorithms have been proposed, including Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017), Reptile (Nichol et al., 2018), and ANIL (Almost No Inner Loop) (Raghu et al., 2019). In this study, we focus on the ANIL, which has demonstrated encouraging outcomes in few-shot classification tasks. ANIL is a meta-learning algorithm that learns to adapt to new tasks by reusing the features learned on previous tasks. The algorithm consists of a feature extractor network and a task-specific classifier. During training, the feature extractor network is updated using a gradient descent step, while the task-specific classifier is updated using an optimization method. The idea is to leverage the learned features to quickly adapt to new tasks with limited data.

Attention is a fundamental mechanism in human cognition that plays a crucial role in filtering and processing information (W. Wang et al., 2019). It allows individuals to selectively focus on relevant aspects of the environment while filtering out irrelevant information. In the context of deep learning, attention mechanisms have been widely used to identify salient features of data that are relevant to the task at hand. The ability of attention mechanisms to capture the most informative parts of the data can lead to more accurate and efficient models. Thus, incorporating attention mechanisms in deep learning algorithms can be a powerful tool for improving model performance and making them more interpretable. Many attention methods have been developed and integrated into various deep neural networks, such as Residual Networks (He et al., 2016) and Convolutional Neural Networks (ConvNets). The attention methods used in this paper include Convolutional Block Attention Module (CBAM) (Woo et al., 2018), Squeeze-and-Excitation Networks (SENet) (Hu et al., 2018), Global Second Order Pooling (GSoP) (Gao et al., 2019), Global Context Network (GC-Net) (Cao et al., 2019), and Efficient Channel Attention Networks (ECA-Net) (Q. Wang et al., 2020). These attention methods aim to improve the image classification accuracy and feature extraction by emphasizing important features and suppressing irrelevant information. By applying these attention methods as plug-ins in deep learning models, we aim to enhance the performance of the model in plant disease classification.

In this study, we evaluate the performance of the ANIL on few-shot classification tasks in the domain of plant recognition, using the ResNet-12 model as the backbone. The ResNet-12 model is pre-trained on the ImageNet dataset (Deng et al., 2009) and used as a feature extractor. In our study, we utilize the attention mechanisms discussed earlier to obtain attended features from the input features. These attended features are fed into the ANIL meta-learning algorithm to train few-shot classifiers with different attention methods. To obtain a robust and accurate classification, we employ ensemble learning called soft voting, which combines the predictions of different classifiers. Finally, we evaluate the performance of the classifiers on a held-out test set and compare their performance. Our evaluation aims to determine the effectiveness of different attention methods for few-shot classification tasks in the agricultural domain.

Related Works

Despite the notable achievements of deep learning in numerous domains (Bayat & Işık, 2022; Gündüz & Işık, 2023; Karaman et al., 2023; Pacal, 2022), its performance remains inadequate in certain areas, such as plant disease recognition, primarily due to the lack of data. Plant disease recognition has been a popular research area in recent years due to its potential impact on crop yields and food security. In (Kaya et al., 2019), they investigate the impact of transfer learning models on deep neural network-based plant classification for four public datasets. The study shows that transfer learning can enhance automated plant identification and improve the performance of low-performing plant classification models. Keçeli et al. (2022) propose a multi-task learning approach for plant species and disease prediction, which has shown to perform better than individual models. The study uses a multi-input neural network that combines raw images and transferred features from a pre-trained deep model, and evaluates the approach on public datasets. Various few-shot learning methods have been proposed to address the challenges of limited data and feature representation in plant disease recognition. In (Argüeso et al., 2020), a siamese network with triplet loss was utilized to recognize six different plant diseases in the Plant Village dataset. Chen et al. (2021) proposed a meta-learning-based method to detect plant diseases of unseen categories with few annotated examples and to identify important input regions for predictions. The authors also contributed a new dataset containing 26 plant species and 60 plant diseases. The use of attention mechanisms in plant disease recognition has also gained popularity in recent years. Lin et al. (2022a) proposed a network that combined cascaded multi-scale features and channel attention to address the challenges of limited features and generalization in few-shot learning for plant disease recognition. Additionally, Lin et al. (2022b) employed the Discrete Cosine Transform (DCT) to convert RGB colors into the frequency domain and extracted important frequency components, which were then processed with a Gaussian-like calibration module to achieve a Gaussian distribution. These works demonstrate the potential of few-shot learning methods, meta-learning, and attention mechanisms in plant disease recognition.

MATERIALS AND METHODS

This section provides an overview of the ANIL algorithm and the attention methods utilized in this study. Additionally, we present the dataset used in our experiments. Finally, we introduce our novel ensemble-based technique for few-shot classification.

Anıl Algorithm

Few-shot learning involves two stages: the initial training stage (meta-training) where a model is trained to be adaptable to new tasks, and the subsequent adaptation stage (meta-testing) where the trained model is adapted to perform new tasks. Meta-learning algorithms are designed to optimize the adaptation algorithm during training, in a way that facilitates learning to learn. Few-shot learning primarily involves classifying rare classes during the meta-testing phase.

During the training stage (for base or seen classes), a model f_θ is learned using a training algorithm on a dataset D^{train}. During the subsequent adaptation stage (for unseen or novel classes), a set of tasks T for few-shot classification is formulated using the test dataset D^{test}, where the classes differ from those in D^{train}. Episodic training is a method employed to accomplish few-shot learning. This approach involves the creation of a set of tasks, T, for few-shot classification by randomly selecting subsets of classes from the pool of previously seen classes. Each task T_i comprises a support set S for model adaptation and a query set Q for model evaluation, with identical label categories to S. In an N-way K-shot task, if there are N categories in the support set S and K examples in each category, the query set Q has N categories and W samples per category. The primary objective is to accurately classify all $N \times W$ samples in Q into their respective N categories.

The core concept of meta-learning is to mimic evaluation scenarios by randomly selecting tasks T from previously seen classes. Specifically, N-way K-shot tasks are generated by selecting a random subset of N classes from the pool of previously seen classes, and then randomly selecting K examples from each of the N classes. The adaptation stage leverages the trained model f_θ and the support set S as inputs, and generates a new classifier. The generated classifier will be assessed on the query set Q to determine its ability to generalize.

In the model-agnostic meta-learning (MAML) (Finn et al., 2017), two types of parameter updates are performed, namely, the outer loop and the inner loop. The outer loop updates the initial parameters of the neural network to facilitate rapid adaptation to new tasks, whereas the inner loop performs task-specific adaptation using a few labeled samples. In 2019, Raghu et al. (2019) proposed ANIL (almost no inner loop), a simplified version of the MAML algorithm, to test the hypothesis that the rapid learning performance of MAML can be achieved solely through feature reuse, as shown in Figure 1. The ANIL algorithm is computationally faster than MAML and has been shown to be equally effective. The ANIL algorithm aims to address the computational inefficiency of the inner loop optimization process in MAML. ANIL reduces the computational cost of MAML, resulting in a simpler and faster training process.

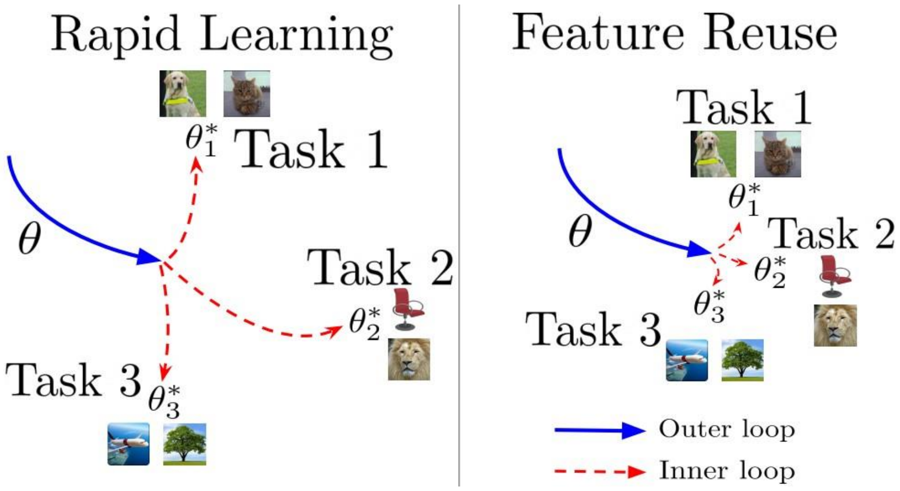

Figure 1. The ANIL algorithm is characterized by its ability to rapidly learn and reuse features for few-shot classification tasks (Raghu et al., 2019). It achieves this by leveraging its meta-learning capability to learn a set of initial parameters that can quickly adapt to new tasks

The main idea behind ANIL is to approximate the inner loop optimization process by only taking one or a few gradient steps on the support set. This is achieved by computing the gradients of the loss function with respect to the model parameters on the support set and then updating the model parameters using these gradients. The updated model parameters are then used to compute the loss function on the query set. After updating the model parameters on the support set, the loss function is computed on the query set using the updated parameters. The gradients of the loss function on the query set with respect to the initial model parameters are then computed and the model parameters are updated using these gradients. The algorithmic representation for the ANIL approach is as follows:

Algorithm 1. ANIL approach

```
1. Initialize model parameters with meta-initialization
2. For each task in training set:
2.1 Compute loss on small batch of examples using only head layer
2.2 Compute gradients of loss with respect to head layer parameters
2.3 Update head layer parameters using gradients
3. Remove head layer after training
4. Use learned representations for unseen tasks without adaptation
```

Attention Methods

In this study, five distinct attention modules were utilized. A brief overview of these modules will be presented in this section.

The first attention-based method we used is SENet, which was proposed by (Hu et al., 2018). The key feature of SENet is the integration of a squeeze-and-excitation (SE) module within the network architecture. The module adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels. This makes it a well-suited for tasks where important features may vary across channels. The excitation operation learns to selectively emphasize informative channels and suppress less useful ones, leading to improved accuracy and efficiency in various computer vision tasks. SENet has achieved state-of-the-art performance on several image classification benchmarks, including ImageNet, CIFAR, and Pascal VOC (Hu et al., 2018). In this study, we applied the SE module to the features extracted from a ResNet-12 model. The authors of SENet have demonstrated that the method is generalizable and can be applied to various pre-existing network architectures, including ResNet and Inception (Szegedy et al., 2016). By incorporating the SE module, we aimed to enhance the discriminative power of the ResNet-12 features and improve the performance of the few-shot classification task.

The other popular attention mechanism for deep neural networks that we used is CBAM, which was proposed by (Woo et al., 2018). CBAM aims to selectively highlight important features of an input image by integrating both channel and spatial attention mechanisms. This makes it a good choice for improving classification performance in situations where feature selection is crucial. The channel attention mechanism uses a SE module, similar to the SENet, to reweight the importance of each feature map. The spatial attention mechanism then refines the feature maps by selectively attending to informative regions within the feature maps. CBAM has shown promising results in various computer vision tasks and can be easily integrated with pre-existing network architectures (Woo et al., 2018). In our study, we used the CBAM module after obtaining features from the ResNet-12 model and evaluated its performance.

GC-Net is the third attention-based method we used in this study, proposed by (Cao et al., 2019). The key idea of GC-Net is to model the global contextual information by performing a weighted average over the feature maps of an entire image, which is then used to reweight the original features. This mechanism has been shown to be effective in various visual recognition tasks, including image classification and segmentation. Global context modeling is achieved through a global pooling layer followed by a two-layer fully connected network. The authors demonstrated that GC-Net can effectively capture long-range dependencies and contextual information in images. GC-Net can be integrated into pre-existing network architectures, such as ResNet, to improve their performance. In our study, we also employed GC-Net as an attention mechanism to enhance the features obtained from the ResNet-12 model.

GSoP is a convolutional network-based method proposed by (Gao et al., 2019) that leverages the second-order statistics of features to enhance the performance of deep neural networks. This approach is based on computing the covariance matrix of feature maps and then using it to perform global pooling, which results in a compact, informative representation of the feature maps. GSoP aggregates second-order statistics of the feature maps in CNNs. This pooling method can capture higher-order statistical dependencies between the channels and the spatial locations of the feature maps. The authors demonstrated that GSoP can be used as a plug-in module for different types of deep neural networks, including ResNet, and can significantly improve their performance on various computer vision tasks, such as image classification, object detection, and semantic segmentation. In this study, we employed GSoP as an attention method after the ResNet-12 model and evaluated its performance for few-shot classification.

ECA-Net is the last attention-based method we used, proposed by (Q. Wang et al., 2020) that efficiently models interdependencies between channels of convolutional features. ECA-Net efficiently models channel dependencies by introducing a novel 1D convolutional operation that adaptively recalibrates feature maps based on channel-wise statistics. The authors demonstrated that their approach achieves state-of-the-art performance on several image classification benchmarks, including ImageNet, with fewer parameters and computational overhead than competing methods such as SENet. In our study, we applied ECA-Net as an attention mechanism on top of features extracted from the ResNet-12 model.

The attention modules utilized in this study implement unique attention mechanisms and concentrate on distinct aspects of the image, such as channel or spatial features. These differences prompted us to integrate all attention modules to construct a robust model. Figure 2 presents the attention modules employed in this study, which consist of several convolutional and computational operations to obtain attended features. The attention methods discussed in Table 1 are widely used in various computer vision applications, including classification and detection. The table provides details on the attention mechanism used, the range of attention weights, and the type of information utilized in the image. Soft attention is a method (Shen et al., 2018) in which the model generates a weight for each input feature, and then takes a weighted sum of the features to obtain a context vector. The weights are generated through a softmax function, which ensures that they sum up to one. Hard attention, on the other hand, is a method in which the model explicitly selects a single feature to focus on. This is typically done through a discrete sampling process, such as using the argmax operation. All attention methods utilized in this study are soft attention-based methods.

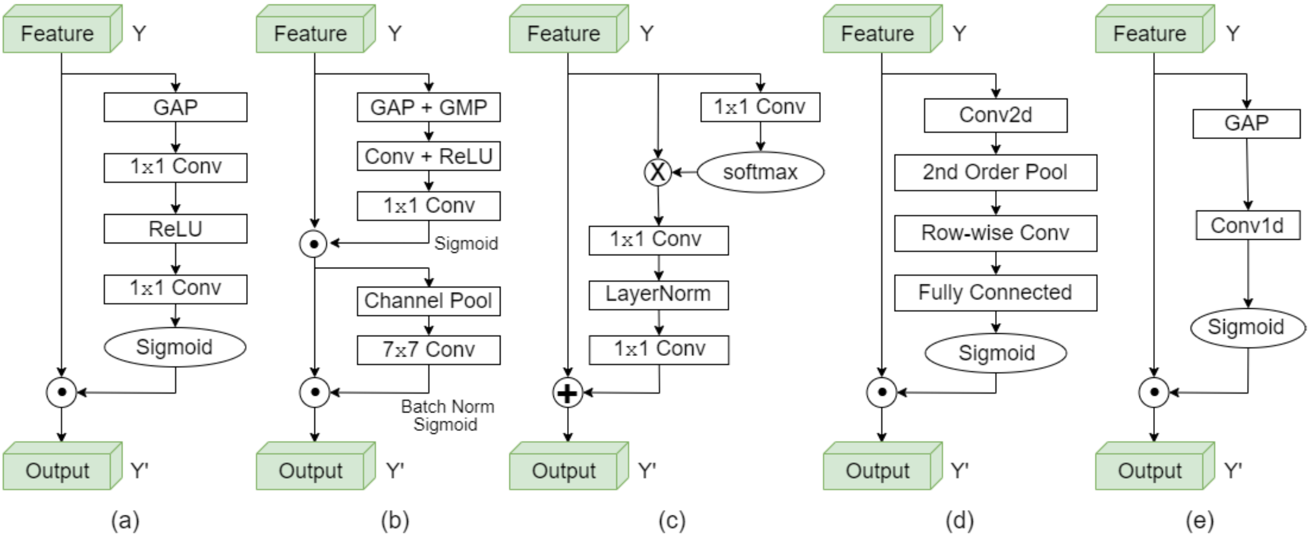

Figure 2. The attention mechanisms used (a) SE Module (b) CBAM Module (c) GC Module (d) GSoP Module (e) ECA Module. ⊙ broadcast element-wise multiplication, ⊕ represents broadcast element-wise addition, and ⊗ denotes matrix multiplication. GAP: global average pooling, GMP: global max pooling. The variable Y represents the obtained features, while Y' represents the attended features.

Table 1. General characteristics of the attention methods. Some of the information presented in this table was sourced from (Guo et al., 2022)

Attention method	Tasks	Attention	Ranges	Soft or Hard	Channel or Spatial	Goal
SE-Net (Hu et al., 2018)	Classification Detection	Channel-wise prod.	(0,1)	Soft	Channel	What to attend
CBAM (Woo et al., 2018)	Classification Detection	Element-wise prod.	(0,1)	Soft	Channel & Spatial	What & where to attend
GC-Net (Cao et al., 2019)	Classification Detection Instance seg.	Self-attention	(0,1)	Soft	Spatial	Where to attend
GSoP (Gao et al., 2019)	Classification	Channel-wise prod.	(0,1)	Soft	Channel	What to attend
ECA-Net (Q. Wang et al., 2020)	Classification Detection	Channel-wise prod.	(0,1)	Soft	Channel	What to attend

Dataset

The effectiveness of our proposed approach was evaluated using the widely recognized Plant Village dataset introduced by (Mohanty et al.) in 2016. The dataset consists of 54306 leaf images belonging to 14 different plant species, which are further divided into 38 classes, with 26 unhealthy and 12 healthy classes. The images in the dataset are colorized and measure 256x256 pixels in size, with significant class imbalance in terms of sample numbers. To address this issue, we employed data augmentation techniques, which involved increasing the number of samples in classes with fewer than the average number of samples to 1500, while randomly selecting 1500 samples from classes with more than the average number of samples. This method resulted in a balanced distribution of samples across all classes. Additionally, we selected the tomato classes to be used in our experiments, which are shown in Figure 3. In the meta-training stage, we utilized a set of 28 base classes, while in the meta-testing stage, we employed 10 classes of tomatoes as the target classes. We chose to use tomato-related classes to facilitate comparisons with other studies. This decision was made with the aim of ensuring consistency in evaluating the performance of our proposed approach.

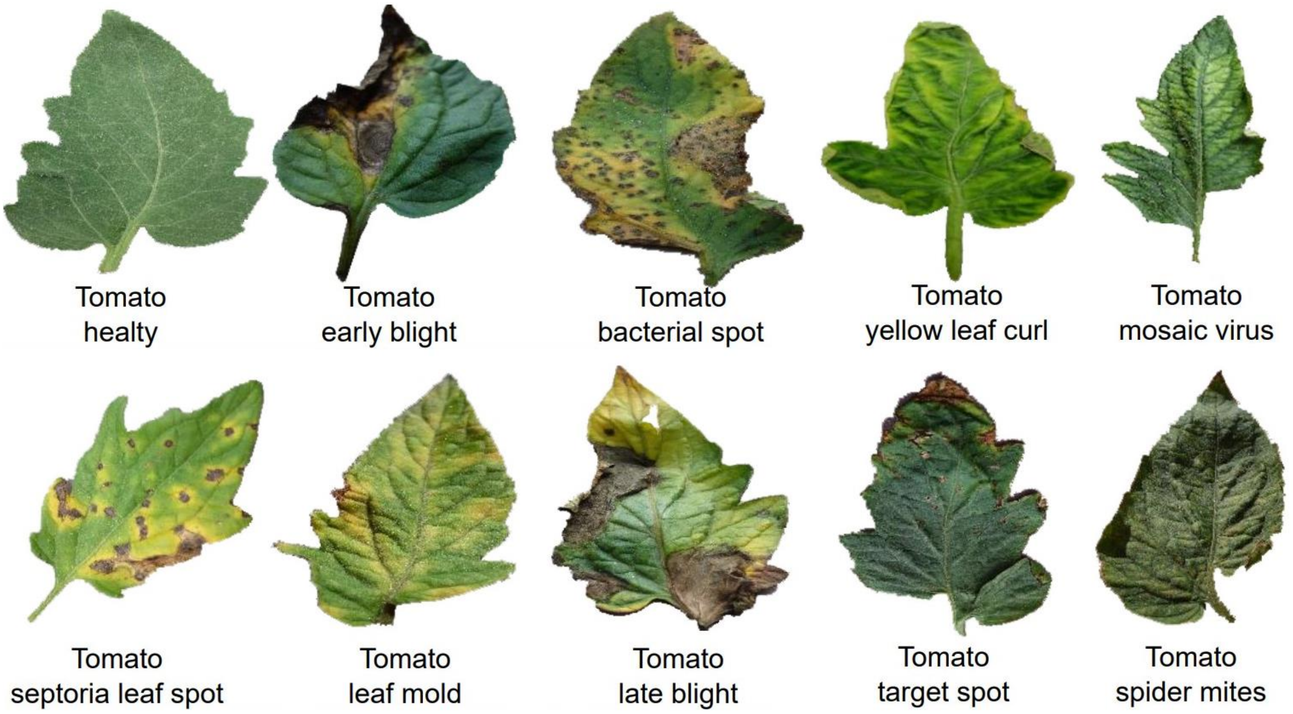

Figure 3. In the meta-testing stage, a total of 10 tomato classes were employed as target classes

Proposed Approach

Our proposed approach is capable of adjusting its attention during the training process. The general overview of the network structure and the training process of the model are described in detail below.

In few-shot learning, feature extraction (f_θ) is an essential step in building a model that can recognize novel classes with only a few examples. One common approach is to use a backbone network for feature extraction, which extracts informative features from the input images. In this study, we utilized ResNet-12 (He et al., 2016), a well-known deep convolutional neural network architecture that has been pre-trained on the large-scale ImageNet dataset (Deng et al., 2009). By utilizing this backbone network, we were able to extract rich and informative features from all plant images, which can be used to train a few-shot learning model.

When the ResNet-12 network is used as the backbone, the initialization of the meta-learning algorithm is achieved by using the embedding component of the pre-trained model after removing the classifier. A widely used approach in computer vision is to extract features from the last fully connected layer before the softmax, as described by (Vinyals et al., 2016). This layer is typically referred to as the *embedding layer* and it outputs a fixed-length feature vector for each input image. This feature vector can be used as a representation of the input image, which can then be used for various tasks. The size of the feature vector is typically determined by the number of neurons in the embedding layer, and in ResNet-12, it has a size of 512. These features were then processed further using attention mechanisms to enable the model to focus on the most relevant information for each task during meta-learning.

In our few-shot learning approach, we used five attention modules to process the ResNet-12 features. As shown in Figure 2, the attention modules consisted of convolutional and computational operations to obtain the attended features. To obtain attended features of the same dimension as the input features, we used element-wise multiplication broadcasting ($\odot$) between the input features and the output of the attention operations. This allowed us to regulate the attention during the training process and improve the performance of the model in few-shot learning. The attended features obtained separately from the five attention modules were flattened and fed into a multi-layer perceptron (MLP, Classifier i) for label prediction. The cross-entropy loss function was used to calculate the prediction errors, and the ADAM optimizer was used to minimize these errors during training.

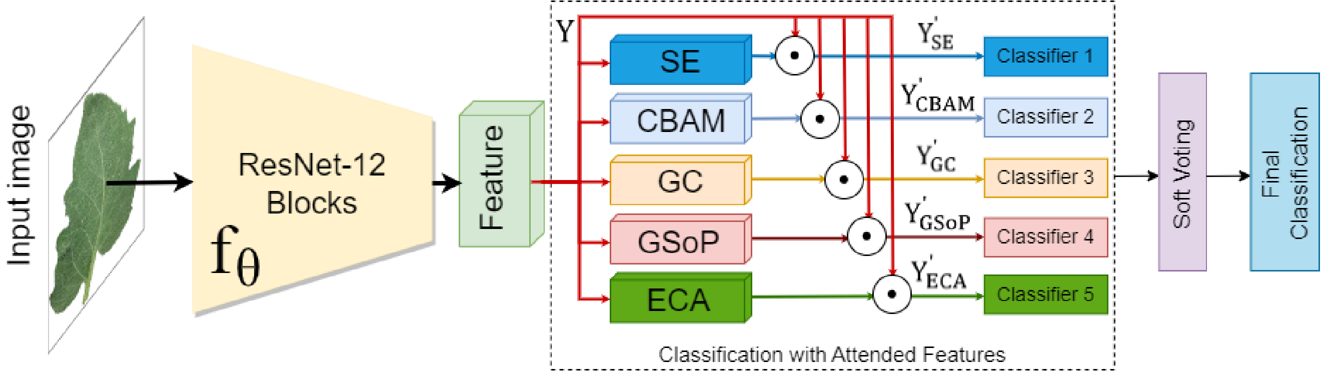

Figure 4. General overview of our proposed method. We utilized features obtained from various attention modules to obtain attended features. These attended features were then used for few-shot classification. To determine the final prediction, we employed a soft voting technique, which assigns weights to the outputs of each classifier. $\odot$ denotes element-wise multiplication

In most cases, the attention mechanism is plugged into the residual blocks, which requires retraining the ResNet model on the ImageNet dataset. Therefore, using a pre-trained model would not be advantageous in this scenario. In this study, attention mechanisms were used after obtaining ResNet features, thereby avoiding the need to retrain the ResNet model on Imagenet. By doing so, the ResNet model was used only once to extract features, and this prevented the need to train ResNet on Imagenet and use it as a feature extractor for five attention models separately. The proposed architecture saved time and processing power, which enabled the testing of models quickly.

Soft voting is a technique that involves combining the outputs of multiple classifiers to perform the final classification. In this approach, the outputs of each classifier are considered as a probability distribution over the possible classes. The final predicted class is then determined by taking the weighted average of the probability distributions of each classifier. This approach can improve the overall classification performance by leveraging the strengths of each classifier and compensating for their weaknesses. In our study, we used soft voting to combine the outputs of the five classifiers obtained from the attended features of the ResNet-12 network. This allowed us to take advantage of the diversity of the attention modules and obtain more accurate predictions. The soft voting technique has been shown to be effective in various applications and is widely used in ensemble learning. The soft voting formula is represented as follows:

$$\hat{y} = \arg\max_{i} \sum_{j=1}^{m} w_j p_{ij} \tag{1}$$

The index i denotes the possible class labels, while the weight assigned to the j-th classifier in the ensemble is represented by w_j. The variable p represents the predicted probabilities for each class, while $\hat{y}$ denotes the final prediction.

RESULTS AND DISCUSSION

This section begins with a presentation of key experimental details. Then, we perform a comparison between our proposed few-shot learning technique and existing state-of-the-art methods for tomato disease detection.

Experimental Details

In our experiments, we selected the set of 10 tomato classes from the Plant Village dataset as the unseen or novel set. The remaining classes of the dataset were utilized as the seen set in the experiments. To ensure consistency, we resize all images in the dataset to a resolution of 84×84 before they are fed into the network.

We utilized the PyTorch implementation of the ANIL meta-learning algorithm through the learn2learn library (Arnold et al., 2020). In the conventional N-way K-shot scenario, the aim is to train a meta-learning model that minimizes the N-way classification loss. To accomplish this objective, we generated several tasks or episodes from the training data within the seen set. Our approach was evaluated on 10-way K-shot ($K = 1, 5, 10, 20$) using a query set of 15 images per class ($W = 15$). To maintain consistency, we used the same task configuration during both the meta-training and meta-testing stages. It is a requirement in meta-learning that the conditions of meta-training and meta-testing match. Specifically, this means that the values of N (10) and W (15) used in meta-training must be identical to those used in meta-testing. This ensures consistency and validity in the evaluation of the meta-learning model's performance.

Deep learning methods commonly rely on a large number of examples to train a model. Meta-learning, on the other hand, involves training a model using thousands of tasks, each consisting of a small number of examples, to improve its ability to learn quickly and effectively. In our case, we trained the meta-learning model using 50,000 randomly generated tasks over 5,000 iterations, with a learning rate of 0.01 that decays to 0.001 after 1,000 iterations. To evaluate our method, we used 600 tasks randomly selected from the unseen classes of the PlantVillage dataset, which is the standard method used in the literature (Dumoulin et al., 2021). The reported accuracies are the average of the accuracies obtained on these 600 tasks, along with their 95% confidence intervals.

Results

This section presents the experimental results obtained through a multi-step process. Firstly, the features of the input images were extracted by leveraging the ResNet-12 backbone network. Next, the attended features were obtained through five distinct attention modules applied to these features. The attended features were then classified using the ANIL meta-learning algorithm in a few-shot learning scenario, utilizing five classifiers separately. Subsequently, a soft voting method was applied to combine the results obtained from these classifiers. Soft voting is an ensemble learning technique that enables input classification by assigning weights to the outputs of individual classifiers.

During the meta-testing stage, we performed few-shot classification on 10 different tomato disease classes, using support sets with different sample sizes (K). It is widely accepted that classification accuracy tends to increase as the sample size in the support set grows larger. Our results are consistent with this expectation, as we observed the greatest increase in accuracy between 1-shot and 5-shot, with all models exhibiting an improvement of around 12 points within this range. As the sample size in the support set increases, the magnitude of the improvement decreases.

As shown in Table 2, the highest accuracy rate was achieved by the model utilizing the CBAM module for all K values. Specifically, the accuracy of the 20-shot model was 97.15%. The CBAM module is designed to process both channel and spatial features of input images, effectively identifying the most distinctive features and their spatial locations. This likely contributes to its superior performance compared to other attention modules. Notably, accurately classifying the 10 different tomato diseases is a challenging task due to the fine-grained nature of the classes, where distinguishing features only vary slightly depending on the disease present on the leaf. In contrast, a coarse-grained classification task such as distinguishing between tomato and apple would be comparatively easier. Following CBAM, the ECA module yielded the most successful results among the other attention modules.

Table 2. Accuracies (%) of the 10-way K-shot classification

Attention + Classifier	1-shot	5-shot	10-shot	20-shot
SE + Classifier 1	80.74	93.27	95.11	95.86
CBAM + Classifier 2	81.12	94.35	96.47	97.15
GC + Classifier 3	80.75	93.10	95.01	95.77
GSoP + Classifier 4	81.05	93.30	95.35	96.12
ECA + Classifier 5	80.98	92.85	95.44	96.23
Soft Voting	**82.20**	**94.88**	**97.05**	**97.66**

The accuracy rates we obtained in classifying the classifier outputs with the soft voting method are quite satisfactory. Specifically, we were able to obtain accuracy rates of 82.20% for 1-shot and 97.66% for 20-shot classification. When compared to previous studies in the literature, it was observed that (Lin et al., 2022b) achieved better results in 1-shot and 5-shot classification. They employed the Discrete Cosine Transform (DCT) to convert RGB colors into the frequency domain and extracted important frequency components, which were then processed with a Gaussian-like calibration module to achieve a Gaussian distribution. Our method, on the other hand, produced better results as the K number increased, as evidenced by Table 3. The accuracy rates achieved with the soft voting method for 10-shot and 20-shot classification were 97.05% and 97.66%, respectively, which are considered state-of-the-art results.

Table 3. Comparison with the literature

Study	Method	1-shot	5-shot	10-shot	20-shot
(Li & Chao, 2021)	Semi-supervised FSL	75.10	90.00	92.70	93.90
(Lin et al., 2022b)	Frequency + Gauss Cal.	**86.34**	**95.30**	96.93	97.48
Ours	Soft Voting	82.20	94.88	**97.05**	**97.66**

Discussion

The results of this study demonstrate that the ANIL algorithm, combined with attention mechanisms, can effectively address the challenges of few-shot learning. The soft voting method used in this study was shown to achieve state-of-the-art results for tomato disease classification. Our results suggest that attention mechanisms can improve the performance of few-shot learning algorithms by enabling more effective feature extraction from limited data. The results also highlight the potential of the ANIL algorithm for addressing the challenges of few-shot learning in other domains.

Furthermore, our study highlights the potential and superiority of our proposed method for fine-grained classification tasks. Specifically, we found that the CBAM module, which processes both channel and spatial features, outperformed other attention modules in this context. Our proposed method offers a promising solution to the challenges of fine-grained classification, where the distinguishing features are often subtle and change only according to the disease on the leaf. However, further research is needed to determine whether this approach can be generalized to other fine-grained classification problems.

One limitation of this study is that we focused exclusively on tomato disease classification. Future research could explore the application of the ANIL algorithm and attention mechanisms to other domains, such as animal or human disease classification. Additionally, future studies could investigate other methods of combining the outputs of multiple classifiers, such as boosting or stacking.

Our study contributes to the growing body of literature on few-shot learning and highlights the potential of attention mechanisms and the ANIL algorithm for addressing challenges in this area. We believe that our proposed approach can be extended to other domains and datasets, and we hope to further exploration of these topics in future research.

CONCLUSION

Deep learning algorithms are known to achieve exceptional results when trained with large amounts of data. However, traditional deep learning methods do not perform well when trained with only a few samples. To address this issue, researchers have developed few-shot learning algorithms, with meta-learning being one such approach. The *Almost No Inner Loop* (ANIL) algorithm is a type of gradient-based meta-learning method. In this study, we utilized the Plant Village dataset to extract features of different tomato classes. We then employed five separate attention modules to obtain attended features from these extracted features. We subsequently classified the attended features using a classifier by flattening them. We separately trained and tested these classifiers with the ANIL algorithm. We used the soft voting method to weigh the classification scores obtained from each classifier and made the final prediction. Our results showed that the soft voting method achieved accuracy rates of 97.05% and 97.66% for 10-shot and 20-shot classification, respectively, which are the state-of-the-art results to the best of our knowledge. Our approach offers a novel way of incorporating attention mechanisms into the feature extraction process and provides a potential avenue for further research into few-shot learning methods.

In future work, we plan to investigate the impact of incorporating additional attention mechanisms into our proposed method to further improve its few-shot learning performance. Additionally, we intend to explore the use of alternative meta-learning algorithms and evaluate their effectiveness in combination with different attention modules.

Conflict of Interest

There is no conflict of interest.

REFERENCES

Albattah, W., Nawaz, M., Javed, A., Masood, M., & Albahli, S. (2022). A novel deep learning method for detection and classification of plant diseases. *Complex & Intelligent Systems*, 1–18.

Argüeso, D., Picon, A., Irusta, U., Medela, A., San-Emeterio, M. G., Bereciartua, A., & Alvarez-Gila, A. (2020). Few-Shot Learning approach for plant disease classification using images taken in the field. *Computers and Electronics in Agriculture*, *175*, 105542.

Arnold, S. M. R., Mahajan, P., Datta, D., Bunner, I., & Zarkias, K. S. (2020). *learn2learn: A Library for Meta-Learning Research*. http://arxiv.org/abs/2008.12284

Bayat, S., & Işık, G. (2022). Recognition of Aras Bird Species From Their Voices With Deep Learning Methods. *Journal of the Institute of Science and Technology*, *12*(3), 1250–1263.

Cao, Y., Xu, J., Lin, S., Wei, F., & Hu, H. (2019). Gcnet: Non-local networks meet squeeze-excitation networks and beyond. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 0.

Chen, L., Cui, X., & Li, W. (2021). Meta-learning for few-shot plant disease detection. *Foods, 10*(10), 2441.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Dumoulin, V., Houlsby, N., Evci, U., Zhai, X., Goroshin, R., Gelly, S., & Larochelle, H. (2021). Comparing transfer and meta learning approaches on a unified few-shot classification benchmark. *ArXiv Preprint ArXiv:2104.02638*.

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. *International Conference on Machine Learning*, 1126–1135.

Gao, Z., Xie, J., Wang, Q., & Li, P. (2019). Global second-order pooling convolutional networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3024–3033.

Gündüz, M. Ş., & Işık, G. (2023). A new YOLO-based method for social distancing from real-time videos. *Neural Computing and Applications*, 1–11.

Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., Zhang, S.-H., Martin, R. R., Cheng, M.-M., & Hu, S.-M. (2022). Attention mechanisms in computer vision: A survey. *Computational Visual Media*, *8*(3), 331–368.

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

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141.

Karaman, A., Pacal, I., Basturk, A., Akay, B., Nalbantoglu, U., Coskun, S., Sahin, O., & Karaboga, D. (2023). Robust real-time polyp detection system design based on YOLO algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (ABC). *Expert Systems with Applications*, *221*, 119741.

Kaya, A., Keceli, A. S., Catal, C., Yalic, H. Y., Temucin, H., & Tekinerdogan, B. (2019). Analysis of transfer learning for deep neural network based plant classification models. *Computers and Electronics in Agriculture*, *158*, 20–29.

Keceli, A. S., Kaya, A., Catal, C., & Tekinerdogan, B. (2022). Deep learning-based multi-task prediction system for plant disease and species detection. *Ecological Informatics*, *69*, 101679.

Li, Y., & Chao, X. (2021). Semi-supervised few-shot learning approach for plant diseases recognition. *Plant Methods*, *17*, 1–10.

Lin, H., Tse, R., Tang, S.-K., Qiang, Z., & Pau, G. (2022a). Few-shot learning approach with multi-scale feature fusion and attention for plant disease recognition. *Frontiers in Plant Science, 13*.

Lin, H., Tse, R., Tang, S.-K., Qiang, Z., & Pau, G. (2022b). Few-Shot Learning for Plant-Disease Recognition in the Frequency Domain. *Plants*, *11*(21), 2814.

Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2021). Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, *35*(1), 857–876.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, *7*, 1419.

Munkhdalai, T., & Yu, H. (2017). Meta networks. *International Conference on Machine Learning*, 2554–2563.

Nichol, A., Achiam, J., & Schulman, J. (2018). On first-order meta-learning algorithms. *ArXiv Preprint ArXiv:1803.02999*.

Pacal, I. (2022). Deep Learning Approaches for Classification of Breast Cancer in Ultrasound (US) Images. *Journal of the Institute of Science and Technology*, *12*(4), 1917–1927.

Patricio, D. I., & Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture*, *153*, 69–81.

Raghu, A., Raghu, M., Bengio, S., & Vinyals, O. (2019). Rapid learning or feature reuse? towards understanding the effectiveness of maml. *ArXiv Preprint ArXiv:1909.09157*.

Shen, T., Zhou, T., Long, G., Jiang, J., Wang, S., & Zhang, C. (2018). Reinforced self-attention network: a hybrid of hard and soft attention for sequence modeling. *ArXiv Preprint ArXiv:1801.10296*.

Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, *30*.

Sun, Q., Liu, Y., Chua, T.-S., & Schiele, B. (2019). Meta-transfer learning for few-shot learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 403–412.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2818–2826.

Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., & others. (2016). Matching networks for one shot learning. *Advances in Neural Information Processing Systems*, *29*.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11534–11542.

Wang, S., Li, C., Wang, R., Liu, Z., Wang, M., Tan, H., Wu, Y., Liu, X., Sun, H., Yang, R., & others. (2021). Annotation-efficient deep learning for automatic medical image segmentation. *Nature Communications*, *12*(1), 5915.

Wang, W., Song, H., Zhao, S., Shen, J., Zhao, S., Hoi, S. C. H., & Ling, H. (2019). Learning unsupervised video object segmentation through visual attention. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3064–3074.

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19.

Yang, J., Guo, X., Li, Y., Marinello, F., Ercisli, S., & Zhang, Z. (2022). A survey of few-shot learning in smart agriculture: developments, applications, and challenges. *Plant Methods*, *18*(1), 1–12.