



ResNet-50Plus: An Attention-Enhanced Residual Backbone for Few-Shot Chili Leaf Disease Classification with Meta-Learning

Rizal Amegia Saputra¹, Agus Buono¹, Karlisa Priandana^{1*}, Samsuzana Abd Aziz², Muhamad Syukur³

¹ School of Data Science, Mathematics, and Informatics, IPB University, Bogor 16680, Indonesia

² Department of Biological and Agricultural Engineering, Universiti Putra Malaysia, Serdang 43400, Malaysia

³ Department of Agronomy and Horticulture, IPB University, Bogor 16680, Indonesia

Corresponding Author Email: karlisa@apps.ipb.ac.id

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310918>

ABSTRACT

Received: 18 May 2026

Revised: 22 August 2026

Accepted: 2 September 2026

Available online: 30 September 2026

Keywords:

few-shot learning, meta-learning, chili leaf disease classification, ResNet-50Plus, Efficient Channel Attention, feature representation

Few-shot plant disease recognition remains difficult when annotated images are scarce and visually similar symptoms occur across classes. This study proposes ResNet-50Plus, an enhanced residual backbone that replaces rectified linear unit (ReLU) with leaky ReLU (LeakyReLU) and incorporates Efficient Channel Attention (ECA) after the fourth residual stage to improve feature representation for meta-learning-based chili leaf disease classification. A dataset of 773 images collected from 175 chili plants was organized into Healthy, Pest, and Virus classes, with plant-level separation used to form development and test sets. ResNet-50Plus was evaluated within First-Order Model-Agnostic Meta-Learning (FOMAML), Prototypical Networks (ProtoNet), and Relation Networks (RelationNet) under 3-way 1-shot, 5-shot, and 10-shot settings. Experiments were repeated across ten random seeds, and performance was assessed using accuracy, standard deviation (Std), coefficient of variation (CV), paired *t*-tests, confusion matrices, *t*-SNE, and Gradient-weighted Class Activation Mapping (Grad-CAM). In the FOMAML setting, ResNet-50Plus achieved 95.49%, 95.43%, and 95.76% accuracy for 1-shot, 5-shot, and 10-shot classification, respectively, consistently outperforming the original ResNet-50. Ablation results showed that both LeakyReLU and ECA contributed to the observed gains, with their combination producing the strongest performance. The proposed backbone also reduced confusion between visually similar Pest and Virus samples and produced more compact class embeddings. These findings indicate that backbone-level feature refinement can improve the accuracy and stability of few-shot chili leaf disease classification under limited-data conditions.

1. INTRODUCTION

Chili peppers are an important horticultural commodity in Indonesia, but disease and pest infestations often reduce productivity by damaging leaves and lowering yields [1, 2]. With recent technological advances, researchers have developed various deep learning approaches for automated plant disease diagnosis, which have proven effective at extracting complex features from images [3]. Nevertheless, most conventional deep learning models rely heavily on large amounts of labeled data to achieve optimal performance [4]. Meanwhile, the collection and labeling of chili leaf disease data require time, incur costs, and require specialized expertise, resulting in limited availability of high-quality datasets.

Data limitations have driven the development of few-shot learning approaches, which enable models to perform classification using only a small number of samples per class [5]. One widely used approach is meta-learning, which trains models on a variety of tasks so that they can quickly adapt to new data with limited samples [6, 7]. Generally, meta-learning is divided into three main groups: optimization-based, metric-based, and black-box meta-learning methods [8].

Optimization-based approaches focus on adapting model parameters through gradient-based learning, such as First-Order Model-Agnostic Meta-Learning (FOMAML). FOMAML is an extension of MAML that eliminates the need for second-order derivative computations, making it more computationally efficient [9]. Meanwhile, metric-based approaches perform classification using similarity or distance measures in the embedding space, such as Prototypical Networks (ProtoNet) and Relation Networks (RelationNet) [10, 11].

Although meta-learning methods such as ProtoNet [4], MAML [12], and RelationNet [11] demonstrate excellent performance in few-shot learning scenarios, the quality of the backbone's feature representations remains a crucial factor, as it influences the model's ability to discriminate between classes within the embedding space. In FOMAML, the quality of the initial representation affects the effectiveness of gradient-based adaptation [13], whereas in ProtoNet, embeddings lacking discriminative power can lead to inter-class overlap [14].

Several studies have been conducted to improve the quality of feature representations in few-shot learning scenarios. Xu

et al. [15] developed ProtoMark by integrating Markov transition features and statistical features into a Transfer Prototype Network for few-shot intrusion detection. In the chili leaf image domain, Suwarningsih et al. [16] applied MAML with Cut-Out, Mix-Up, and Cut-Mix augmentation strategies to improve classification performance under data-limited conditions; however, this study focused on variety identification based on leaf morphological characteristics and did not explore the development of an attention-based backbone for disease classification. Ren et al. [17] developed an adaptive feature selection method by integrating local Convolutional Neural Network (CNN) representations and global Transformer representations in few-shot image classification. Wang et al. [18] proposed FRCAE via Channel-Wise Attention Expansion to expand channel dimensionality, along with feature regularization to maintain consistency between support and query representations. Meanwhile, Li et al. [19] developed MFEML by integrating SimAM and Efficient Channel Attention (ECA) into a CNN-based meta-feature enhancement module for few-shot bearing fault identification using MAML.

These studies have demonstrated performance improvements; the strategies employed include feature fusion, feature reconstruction, and attention-based feature enhancement. However, exploration into modifying the residual backbone through a combination of activation functions and the application of channel attention remains relatively limited, particularly within the context of episodic meta-learning.

In this study, ResNet-50 serves as the backbone. However, ResNet-50 natively employs the rectified linear unit (ReLU) activation function, which yields a zero gradient for negative inputs; consequently, its use can cause neurons to stop learning [20]. Meanwhile, activation functions play a role in maintaining gradient flow during the learning process [21].

Various alternative activation functions have been developed to address this limitation, including leaky rectified linear unit (LeakyReLU) [22], ELU [23], and Swish [24]. LeakyReLU was selected for this study because it preserves small or negative gradients, thereby mitigating the risk of "dying ReLU" (a condition where ReLU neurons become permanently inactive). In addition to modifying the activation function, this study also applies ECA to perform channel recalibration on the feature representations. However, unlike the study by Li et al. [19], ECA is applied to the 4th layer of ResNet-50.

To address these issues, this study proposes ResNet-50Plus, an evolution of ResNet-50 that replaces the ReLU activation function in residual blocks with LeakyReLU and integrates ECA after the fourth residual layer. To ensure a fair comparison, both ResNet-50 and ResNet-50Plus were initially trained on the same domain-specific dataset using identical training configurations. The resulting pre-training weights were then used to initialize parameters for the meta-learning process. Subsequently, both backbones were evaluated using the same experimental configurations across FOMAML, ProtoNet, and RelationNet.

Thus, the main contributions of this study are as follows:

- Proposes ResNet-50Plus, an evolution of ResNet-50 that incorporates LeakyReLU within residual blocks and integrates ECA to enhance classification performance in few-shot learning scenarios.
- Utilizes pre-trained ResNet-50Plus weights for parameter initialization during the meta-learning process to facilitate

model adaptation under data-constrained conditions.

•Evaluates ResNet-50Plus within FOMAML, ProtoNet, and RelationNet frameworks to analyze the consistency of the backbone's performance across different meta-learning approaches and varying numbers of support samples.

The remainder of this paper is organized as follows: Section 2 presents a review of related work. Section 3 discusses the architecture and methodology of the proposed model. Section 4 details the experimental results and model evaluation, while Section 5 concludes the paper.

2. THEORETICAL BACKGROUND

2.1 Few-shot learning

During the training process, few-shot learning generally employs episodic meta-learning, wherein the model is trained through a series of episodes [25]. Each episode is a meta-task τ_i . This process consists of two components: a support set D_i^S and a query set D_i^Q . In the i -th episode, the support set is used as a learning reference and is represented as $S = \{s_1, s_2, \dots, s_n\}$. Meanwhile, a query set is used to evaluate a model's generalization ability and is represented as $Q = \{q_1, q_2, \dots, q_m\}$. Each element within the support set and query set is defined as a data-label pair specifically $s_j = (x_j, y_j)$ and $q_k = (x_k, y_k)$, where x denotes the input data and y denotes the class label. Task configurations in few-shot learning are generally expressed in the form N -way K -shot, indicating that each episode consists of N classes, each with K labeled samples in the support set, so the size support set expressed as $|D_i^S| = N \times K$. Meanwhile, the query set size is expressed as $|D_i^Q| = N \times Q$ [26].

2.2 Meta-learning framework

2.2.1 Model-Agnostic Meta-Learning

MAML [9] is a gradient-based meta-learning approach designed to learn the parameter initialization θ of a model so that it can adapt quickly to new tasks even with limited data.

MAML consists of two optimization processes: the inner loop and the outer loop [27]. In the inner loop, the model adapts to a specific task τ_i by performing one or more gradient descent steps based on the task-specific loss $\mathcal{L}_{\tau_i}$. The parameter updates are expressed in Eq. (1) as follows:

$$\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\tau_i}(f_{\theta}) \quad (1)$$

where, α is the learning rate in the inner loop, and θ'_i denotes the parameters that have been adapted for task τ_i .

In the outer loop, the initial parameter θ is updated based on the performance of the adapted model θ'_i on the adjusted query set [4]. The outer loop process is expressed in Eq. (2) as follows:

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\tau_i \sim p(\tau)} \mathcal{L}_{\tau_i}(f_{\theta'_i}) \quad (2)$$

where, β is the meta learning rate and $p(\tau)$ denotes the task distribution.

The FOMAML model is a simplified sub-variant of MAML that eliminates second-order derivatives when calculating

meta-gradients, thereby making computation more efficient [9]. The FOMAML formula is presented in Eq. (3).

$$\theta \leftarrow \theta - \beta \sum_{\tau_i \sim p(\tau)} \nabla_{\theta'_i} \mathcal{L}_{\tau_i} (f_{\theta'_i}) \quad (3)$$

2.2.2 Prototypical Networks

ProtoNet operates within an embedding space, where the classification process is based on the distance between a query sample and the prototype representation of each class [28].

In the N -way K -shot classification scenario, the prototype for the n -th class in the i -th episode (τ_i) is calculated as shown in Eq. (4) [4, 29]:

$$c_n = \frac{1}{K} \sum_{(x_j, y_j) \in D_i^S} f_{\theta}(x_j) \cdot 1\{y_j = n\} \quad (4)$$

where, c_n is the prototype of the n -th class, D_i^S is the support dataset for the i -th meta-task (τ_i), f_{θ} is an embedding function parameterized by θ , and $1\{y_j = n\}$ is an indicator that equals 1 if $y_j = n$ and 0 otherwise.

In the meta-testing phase, each query sample $x_q \in D_i^Q$ is classified based on its distance from the prototypes. The probability that x_q belongs to class n is calculated using Eq. (5) [30].

$$p_{\theta}(y_q = n | x_q) = \frac{\exp(-d(f_{\theta}(x_q), c_n))}{\sum_{n'=1}^N \exp(-d(f_{\theta}(x_q), c_{n'}))} \quad (5)$$

where, $d(\cdot, \cdot)$ denotes a distance metric, which is typically the Euclidean distance or cosine similarity.

To train the model, the cross-entropy loss function is used, calculated on the query set D_i^Q in each episode. This loss function is expressed by Eq. (6) [4, 30].

$$\mathcal{L} = - \sum_{(x_q, y_q) \in D_i^Q} \log p_{\theta}(y_q | x_q) \quad (6)$$

where, D_i^Q is the query set for the i -th episode, and $\log p_{\theta}(y_q | x_q)$ denotes the logarithm of the predicted probability of the true label for the query sample x_q , where the probability value $p_{\theta}(y_q | x_q)$ is obtained from Eq. (5) by calculating the similarity between the query and the prototype of each class.

2.2.3 Relation Networks

RelationNet [11] is a metric-based method for classification tasks that learns similarity relationships between samples. Unlike ProtoNet, which measures the proximity of samples to class prototypes in an embedding space, this method uses a directly learned relation module to estimate the degree of similarity between support data and query data [4].

The Relation Networks equation is formulated as in Eq. (7):

$$r_{j,q} = g_{\phi}([f_{\theta}(x_j), f_{\theta}(x_q)]) \quad (7)$$

where, $[\cdot, \cdot]$ denotes the concatenation operation, $f_{\theta}(\cdot)$ represents feature extraction, $g_{\phi}(\cdot)$ is the relation module that learns the similarity function, and $r_{j,q}$ is the relation score indicating the degree of similarity between the support samples x_j and x_q the query sample.

2.3 Evaluation

The evaluation metrics used in this study include accuracy, mean, standard deviation (Std), and coefficient of variation (CV), calculated based on test results obtained using ten random seeds (42–51). Additionally, a confusion matrix was constructed by aggregating the results from all tests [31]. Statistical analysis using a paired t -test with a significance level of $p < 0.05$ was performed to evaluate performance differences between models. To enhance model interpretability, the Gradient-weighted Class Activation Mapping (Grad-CAM) technique was employed to visualize the image regions on which the model focused during the classification process [32].

3. PROPOSED METHOD

3.1 Overall framework

This section presents the overall framework of the proposed method. The developed framework integrates a ResNet-50Plus-based residual backbone enhanced with LeakyReLU activation and ECA to improve feature representation quality. In addition to serving as a feature extractor, the pre-trained backbone is used to initialize parameters for the meta-learning process. Furthermore, this backbone is integrated into the FOMAML, ProtoNet, and RelationNet frameworks. An overview of the proposed method is shown in Figure 1.

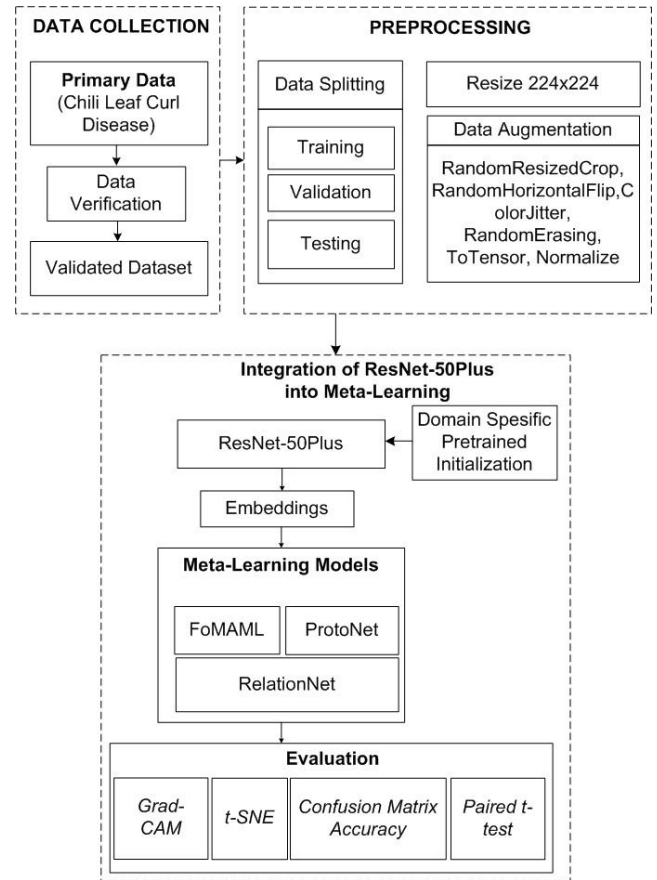


Figure 1. Proposed method

3.2 Data acquisition and preprocessing

The research dataset was obtained from chili plants

cultivated under two environmental conditions: a greenhouse and an open field. The plants were classified into three categories: Healthy, Pest-infested, and Virus-infected. The Healthy category comprised asymptomatic plants grown in a greenhouse, isolated from inoculated plants to minimize the risk of viral exposure. The Virus category included plants inoculated with Pepper yellow leaf curl Indonesia virus (PepYLCIV), a member of the Begomovirus genus with infections confirmed via Polymerase Chain Reaction (PCR) in collaboration with plant pathologists. Identification utilized universal Begomovirus primers SPG1/SPG2 targeting DNA-A (yielding an amplification product of approximately 912 bp [33]), and PepYLCIV-specific DNA-B primers (yielding an amplification product of approximately 385 bp [34]). The PCR results showed positive amplification of DNA-B in symptomatic plants that served as the image sources for the Virus category. Meanwhile, the Pest category consisted of plants exhibiting leaf-curling symptoms caused by thrips infestations in the open field; this status was verified by agricultural experts based on symptom characteristics observed in the field.

Images were captured using a 50 MP Samsung Galaxy A55 camera with automatic settings. Photography took place between 07:30 and 10:00 AM from various angles and positions. Images were recorded at a native resolution of 4080 × 3060 pixels, with lighting and background variations reflecting the environmental conditions at the time of capture. The data collection process yielded 773 images: 260 images of healthy plants, 253 images of pest-infested plants, and 260 images of virus-infected plants. The dataset includes 218 single-leaf images and 555 compound-leaf images. The details are as follows: the Healthy category consists of 55 single-leaf images and 205 compound-leaf images; the Pest category comprises 111 single-leaf images and 142 compound-leaf images; and the Virus category includes 52 single-leaf images and 208 compound-leaf images. This composition provides variation in terms of number of leaves, shooting angles, object positioning, lighting, and backgrounds. Examples of original images from each of the Healthy, Pest, and Virus are shown in Figure 2.

A total of 175 chili plants were used as the data source in this study. From this group, 35 plants were manually selected for testing and yielded 155 images, consisting of 52 healthy, 51 pest-infested, and 52 virus-infected. Meanwhile, the remaining 140 plants produced 618 images, which were subsequently divided into training and validation sets at an 80:20 ratio for each class using a fixed random seed. This split resulted in 493 training images (166 healthy, 161 pest-infested, and 166 virus-infected) and 125 validation images (42 healthy, 41 pest-infested, and 42 virus-infected).

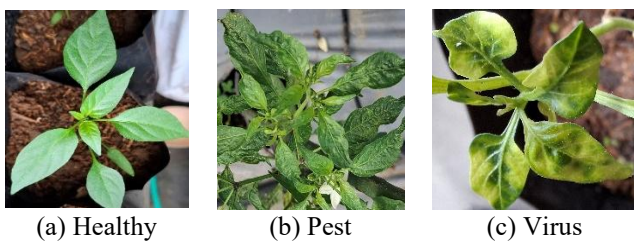


Figure 2. Types of chili leaf conditions

During the preprocessing stage, all images were automatically resized to 224 × 224 pixels to match the backbone’s input dimensions. The training data was

augmented with random horizontal flipping ($p = 0.5$) and color jitter ($p = 0.3$), then converted to tensors and normalized using standard ImageNet parameters. For the validation and test data, only resizing, tensor conversion, and normalization were applied, without random augmentation.

3.3 Backbone network

In this study, ResNet-50 was modified into ResNet-50Plus by replacing all ReLU activation functions with LeakyReLU to improve gradient propagation and mitigate the “dying ReLU” problem [21]. The LeakyReLU function is defined as:

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases} \quad (8)$$

where, α is a small coefficient in the negative region.

In addition, the ECA module is integrated after the fourth residual layer to adaptively enhance channel attention. Unlike conventional attention mechanisms that employ a fixed kernel size, ECA determines the kernel size based on the number of feature map channels using the equation [35]:

$$k = \psi(C) = \lfloor \frac{\log_2(C)}{\gamma} + b \rfloor_{\text{odd}} \quad (9)$$

where, C is the number of feature map channels; γ and b are control constants; $\lfloor \cdot \rfloor_{\text{odd}}$ denotes the nearest odd number.

In this study, the ECA module was applied to the fourth residual layer with a channel count of $C = 2048$, resulting in a kernel size of $k = 7$. The proposed ResNet-50Plus architecture is shown in Figure 3, while modifications to the original ResNet-50 architecture are indicated by dashed red lines.

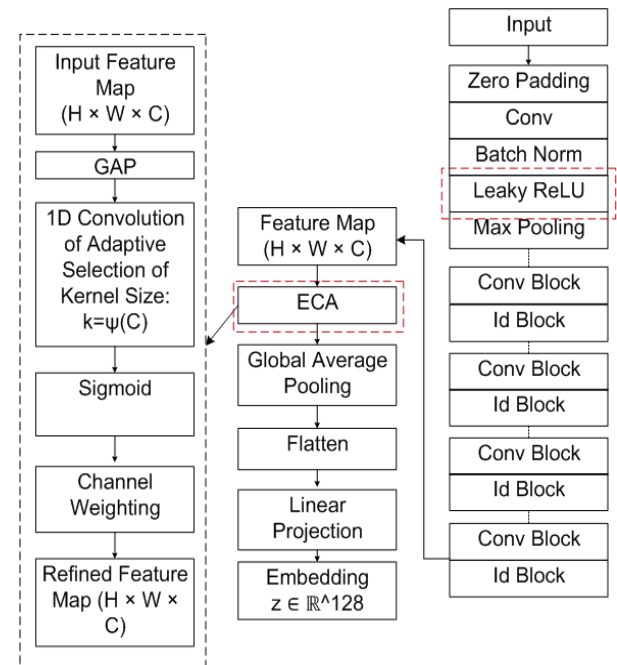


Figure 3. Backbone architecture

4. RESULT AND ANALYSIS

4.1 Implementation and training settings

The experiments were conducted on a Kaggle Notebook

using an NVIDIA Tesla P100 GPU (16 GB VRAM) within a Linux (Ubuntu) environment. The implementation was developed using Python 3.10 with PyTorch, Torchvision, and Scikit-learn libraries. The embedding dimension was set to 128 for all experiments. To ensure reproducibility and result consistency, all experiments utilized controlled random seeds ranging from 42 to 51.

To ensure a fair comparison, ResNet-50 and ResNet-50Plus were pretrained on the same chili leaf disease dataset under identical conditions, including data splitting, augmentation, optimizer, learning rate, batch size, and number of epochs. The resulting pretrained weights served as initialization parameters for all evaluated meta-learning frameworks. Consequently, the only difference between the two backbones lay in the architectural modifications applied to ResNet-50Plus; thus, observed performance differences primarily reflect the impact of the ResNet-50Plus architectural modifications rather than variations in the pretraining process. Details regarding the episodic meta-learning setup are presented in Table 1.

Table 1. Meta-learning settings

Parameter	Setting
Few-Shot Setting	3-way
Shot Variations	1-shot, 5-shot, 10-shot
Query Samples	5 per class
Training Epochs	10
Training Episodes	100 episodes per epoch
Meta-Test Episodes	50
Learning Rate	1×10^{-4}
Inner Learning Rate	1×10^{-2}
Backbone Initialization	Domain-specific pretrained weights (both ResNet-50 and ResNet-50Plus)
Pretraining Dataset	Domain-specific chili leaf dataset
Pretraining Protocol	Identical for ResNet-50 and ResNet-50Plus

4.2 Backbone evaluation and ablation study

To evaluate the impact of backbone modifications and analyze the contribution of each component to meta-learning performance, an evaluation and ablation study was conducted using FOMAML in a 3-way few-shot classification scenario with 1-shot, 5-shot, and 10-shot variations. Four backbone configurations were compared: ResNet-50 as the baseline; ResNet-50 + LeakyReLU to evaluate the impact of changing the activation function; ResNet-50 + ECA to evaluate the contribution of channel attention; and ResNet-50Plus, which combines LeakyReLU and ECA. All configurations used the same experimental protocol and were evaluated based on the average of ten experiments with different random seeds.

The results in Table 2 show that both components individually improve accuracy compared to the ResNet-50 baseline, although their contributions differed. The use of LeakyReLU increased accuracy from 92.72% to 94.63% for 1-shot, from 92.71% to 94.76% for 5-shot, and from 92.83% to 94.81% for 10-shot. Meanwhile, the use of ECA resulted in accuracies of 93.64%, 93.80%, and 93.87%, respectively. These results indicate that LeakyReLU provides a greater individual improvement compared to ECA across all shot settings.

Combining both components in ResNet-50Plus yielded the highest accuracy across all scenarios: 95.49%, 95.43%, and 95.76% for the 1-shot, 5-shot, and 10-shot scenarios, respectively. Compared to ResNet-50, the accuracy improvements were 2.69, 2.98, and 2.61 percentage points,

respectively. The results of this ablation study indicate that LeakyReLU and ECA each contribute to improved classification performance, while their integration yields superior performance compared to using either component in isolation.

Table 2. Backbone evaluation and ablation study under First-Order Model-Agnostic Meta-Learning (FOMAML)-based few-shot classification

Backbone Configuration	Shot Setting	Accuracy Mean $\pm$ Std (%)	CV (%)
ResNet-50	1-shot	92.72 ± 1.48	1.60
	5-shot	92.71 ± 1.23	1.32
	10-shot	92.83 ± 1.42	1.53
ResNet-50+LeakyReLU	1-shot	94.63 ± 1.22	1.29
	5-shot	94.76 ± 0.72	0.76
	10-shot	94.81 ± 0.91	0.96
ResNet-50+ECA	1-shot	93.64 ± 1.62	1.73
	5-shot	93.80 ± 1.58	1.69
	10-shot	93.87 ± 1.72	1.84
ResNet-50Plus	1-shot	95.49 ± 1.14	1.19
	5-shot	95.43 ± 1.00	1.05
	10-shot	95.76 ± 1.14	1.19

Note: Std = standard deviation, CV = coefficient of variation, ECA = Efficient Channel Attention.

Table 3. Efficient Channel Attention (ECA) placement evaluation

ECA Placement	Shot Setting	Accuracy Mean $\pm$ Std (%)	CV (%)
Layer 1	1-shot	94.62 ± 0.89	0.94
	5-shot	94.30 ± 1.16	1.23
	10-shot	94.95 ± 0.62	0.65
Layer 2	1-shot	94.67 ± 0.72	0.76
	5-shot	94.55 ± 1.40	1.48
	10-shot	94.99 ± 1.13	1.19
Layer 3	1-shot	94.77 ± 1.37	1.45
	5-shot	94.52 ± 1.86	1.97
	10-shot	95.27 ± 0.95	1.00
Layer 4	1-shot	95.49 ± 1.14	1.19
	5-shot	95.43 ± 1.00	1.05
	10-shot	95.76 ± 1.14	1.19

Note: Std = standard deviation, CV = coefficient of variation.

To provide experimental support for the choice of ECA position, an additional evaluation was conducted comparing ECA placement after the Layer 3 and Layer 4 residual blocks in ResNet-50Plus, using the same experimental protocol. Placing the ECA in Layer 4 consistently yielded higher average accuracy across all shot scenarios, namely 95.49%, 95.43%, and 95.76%, compared to Layer 3, which achieved 94.77%, 94.52%, and 95.27%. Thus, Layer 4 provided improvements of 0.72, 0.91, and 0.49 percentage points for the 1-shot, 5-shot, and 10-shot scenarios, respectively. Regarding performance variability, Layer 4 also produced lower Std and CV values in the 1-shot and 5-shot scenarios. However, in the 10-shot scenario, Layer 3 exhibited slightly lower variability, even though its average accuracy remained below that of Layer 4. These results provide experimental support for selecting Layer 4 as the ECA position in ResNet-50Plus. The complete results of the comparison between ECA placement at Layer 3 and Layer 4 are presented in Table 3.

4.3 Comparative analysis of meta-learning methods

The performance comparison in Table 4 reveals distinct

adaptation characteristics for each meta-learning method. FOMAML consistently achieved the highest average accuracy across all shot variations: 95.49% for 1-shot, 95.43% for 5-shot, and 95.76% for 10-shot. Furthermore, the relatively low Std and CV values indicate that FOMAML maintains stable performance across various random seeds. This reinforces the finding that FOMAML’s gradient-based adaptation mechanism is more effective at generating parameters capable of adapting to diverse few-shot learning episodes.

ProtoNet also demonstrated competitive performance with accuracies of 93.85%, 94.55%, and 94.78% for the 1-shot, 5-shot, and 10-shot scenarios, respectively. Its performance variance tended to decrease as the number of support samples increased; this is evidenced by the drop in Std from 1.43 in the 1-shot scenario to 1.06 in the 10-shot scenario, alongside the lowest CV value of 0.81% in the 10-shot scenario. This indicates that ProtoNet derives increasingly stable prototypical representations as the number of support samples grows. Meanwhile, RelationNet achieved accuracies of 92.67%, 94.14%, and 94.52% across the respective shot scenarios. Although its performance improved with a larger number of support samples, the method still exhibited relatively higher result variance compared to FOMAML, particularly in the 1-shot and 5-shot scenarios, with Std values of 2.48 and 2.03, respectively. This suggests that RelationNet’s similarity-learning-based approach is more sensitive to episode variations and embedding quality than the optimization-based approach employed by FOMAML.

Table 4. Comparison of meta-learning methods in few-shot classification

Method	Shot Setting	Accuracy	CV (%)
		Mean ± Std (%)	
FOMAML	1-shot	95.49 ± 1.14	1.19
	5-shot	95.43 ± 1.00	1.05
	10-shot	95.76 ± 1.14	1.19
ProtoNet	1-shot	93.85 ± 1.43	1.94
	5-shot	94.55 ± 1.46	1.23
	10-shot	94.78 ± 1.06	0.81
RelationNet	1-shot	92.67 ± 2.48	1.13
	5-shot	94.14 ± 2.03	1.30
	10-shot	94.52 ± 1.74	1.64

Note: FOMAML = First-Order Model-Agnostic Meta-Learning, ProtoNet = Prototypical Networks, RelationNet = Relation Networks, Std = standard deviation, CV = coefficient of variation.

4.4 Performance analysis and visualization

4.4.1 Backbone performance comparison

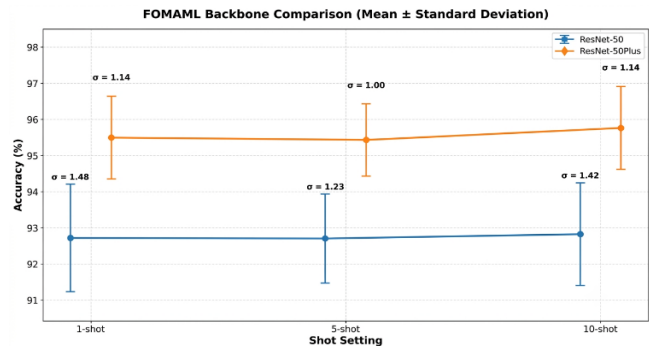


Figure 4. Performance comparison of ResNet-50 and ResNet-50Plus under First-Order Model-Agnostic Meta-Learning (FOMAML) across shot settings

The evaluation results for both backbones indicate that ResNet-50Plus exhibits a relatively stable performance pattern across all shot variations, with minimal accuracy fluctuations. Although there was a slight drop in accuracy in the 5-shot scenario, performance improved again in the 10-shot scenario, while the Std remained low. Furthermore, the error bars for ResNet-50Plus tended to be smaller than those for ResNet-50. In contrast, ResNet-50 displayed greater Std fluctuations across the various scenarios, making it more sensitive to changes in data distribution between episodes. A visualization of this performance comparison is presented in Figure 4.

4.4.2 Accuracy trends across different shot settings

The visualization in Figure 5 shows that increasing the number of support samples tends to improve accuracy across all meta-learning methods. ProtoNet and Relation Networks demonstrated gradual performance improvements from the 1-shot to the 10-shot setting, whereas FOMAML maintained the highest accuracy across all test scenarios, despite a slight dip at the 5-shot level. Additionally, FOMAML also exhibited relatively smaller error bars in the 1-shot and 5-shot scenarios. In the 10-shot scenario, all methods showed increased accuracy. However, the overall Std values were quite high, particularly for Relation Networks, indicating that the performance of these methods became more sensitive to variations in the number of episodes.

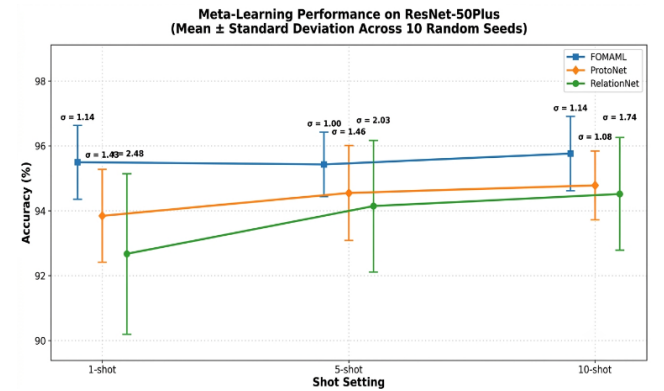


Figure 5. Accuracy comparison of meta-learning methods across shot settings

4.4.3 Statistical significance analysis

A paired *t*-test was conducted to evaluate the significance of differences in backbone performance within the FOMAML framework using seed-level accuracy from 10 random seeds (42–51). Each comparison consisted of 10 pairs of observations ($n = 10$, $df = 9$) matched by seed, with each accuracy value representing the aggregated result of all testing episodes within a single experimental run. The results in Table 5 show that ResNet-50Plus significantly outperformed ResNet-50 across all shot variations ($p < 0.05$), with mean differences of 2.77, 2.72, and 2.93 percentage points in the 1-shot, 5-shot, and 10-shot scenarios, respectively. The largest improvement was observed in the 10-shot scenario, while the highest statistical significance was observed in the 5-shot scenario ($p = 0.000105$).

The results in Table 6 indicate that FOMAML outperformed other meta-learning methods, particularly in the 1-shot scenario. FOMAML significantly outperformed ProtoNet and RelationNet, with mean differences of 1.646 and 2.819 percentage points, respectively ($p = 0.028$ and $p = 0.018$). These results demonstrate the empirical superiority of

FOMAML in conditions where the number of support samples is highly limited. Conversely, in the 5-shot and 10-shot scenarios, no significant differences were found between the methods ($p \geq 0.05$), although FOMAML still achieved the

highest average accuracy. These results indicate that the performance gap between FOMAML, ProtoNet, and RelationNet narrows as the number of support samples increases.

Table 5. Paired t -test results for backbone comparison under First-Order Model-Agnostic Meta-Learning (FOMAML)

Shot Setting	Comparison	Mean Difference (%)	t -Value	p -Value	Significance
1-shot	ResNet-50Plus vs ResNet-50	2.77	3.794	0.004260	*
5-shot	ResNet-50Plus vs ResNet-50	2.72	6.552	0.000105	*
10-shot	ResNet-50Plus vs ResNet-50	2.93	5.156	0.000599	*

Note: * = significant ($p < 0.05$).

Table 6. Paired t -test results for performance comparison across meta-learning methods and shot settings

Shot Setting	Comparison	Mean Difference (%)	t -Value	p -Value	Significance
1-shot	FOMAML vs ProtoNet	1.646	2.611	0.028	*
1-shot	FOMAML vs RelationNet	2.819	2.889	0.018	*
1-shot	ProtoNet vs RelationNet	1.173	1.180	0.268	ns
5-shot	FOMAML vs ProtoNet	0.881	1.229	0.250	ns
5-shot	FOMAML vs RelationNet	1.288	1.563	0.152	ns
5-shot	ProtoNet vs RelationNet	0.407	0.736	0.480	ns
10-shot	FOMAML vs ProtoNet	0.98	1.661	0.131	ns
10-shot	FOMAML vs RelationNet	1.24	1.546	0.156	ns
10-shot	ProtoNet vs RelationNet	0.26	0.581	0.576	ns

Note: ns = not significant ($p \geq 0.05$); * = significant ($p < 0.05$), FOMAML = First-Order Model-Agnostic Meta-Learning, ProtoNet = Prototypical Networks, RelationNet = Relation Networks.

4.4.4 Confusion matrix analysis

To evaluate classification performance at the class level, a confusion matrix was constructed by aggregating the prediction results from all test episodes and seeds within the 10-shot scenario. The evaluation utilized 50 testing episodes per seed, with 3 classes and 5 query samples per class. Across ten different seeds (42–51), the confusion matrix represents a total of 7,500 query samples. All values were subsequently normalized to percentages to facilitate the analysis of prediction distributions for each class. The confusion matrix results are presented in Figures 6 and 7.

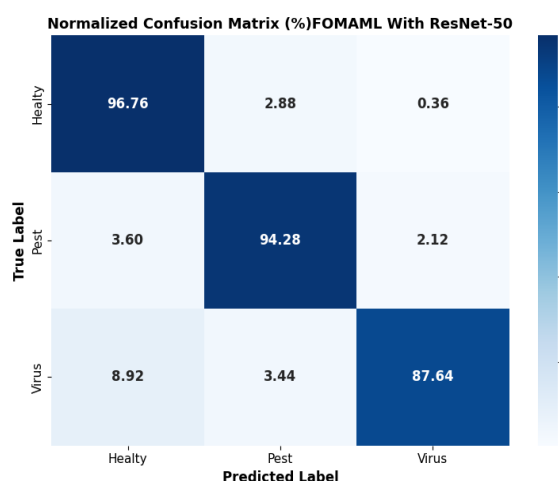


Figure 6. Normalized aggregate confusion matrix of First-Order Model-Agnostic Meta-Learning (FOMAML) with ResNet-50 (10-shot)

Based on Figures 6 and 7, FOMAML with ResNet-50Plus yielded more accurate prediction distributions than ResNet-50 across all classes. The most significant improvement occurred in the Pest class, where accuracy rose from 94.28% to 97.36%. This improvement was accompanied by a reduction in misclassifications into the Healthy class (from 3.60% to

1.16%) and the Virus class (from 2.12% to 1.48%). For the Virus class, accuracy increased from 87.64% to 91.88%, while misclassifications into the Pest class decreased from 3.44% to 2.12%. Similarly, for the Healthy class, accuracy improved from 96.76% to 98.04%, with misclassifications into the Pest class dropping from 2.88% to 1.44%. Overall, these results demonstrate that ResNet-50Plus reduces classification errors across nearly all class pairs, resulting in a more consistent prediction distribution compared to the ResNet-50 backbone. Nevertheless, the Virus class remains the most difficult to classify. As much as 6.00% of the Virus samples were still predicted as Healthy even though all samples in the Virus class had been confirmed via PCR analysis prior to the image acquisition process.

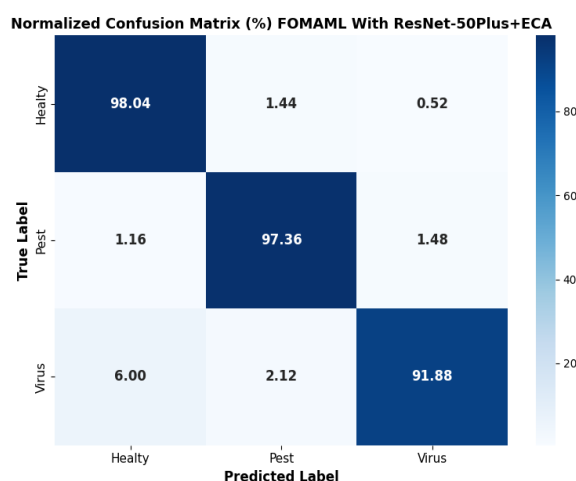


Figure 7. Normalized aggregate confusion matrix of First-Order Model-Agnostic Meta-Learning (FOMAML) with ResNet-50Plus (10-shot)

Therefore, this misclassification is not caused by labeling bias in the dataset but rather reflects the intrinsic visual similarity between healthy leaves and leaves in the early stages

of viral infection. During this initial phase, leaf discoloration and curling patterns have not yet become clearly apparent, resulting in limited available visual information, despite the plants being biologically confirmed as infected. This situation indicates that the main challenge in classifying chili leaf curl disease does not lie in the validity of the labels but rather in the high degree of visual similarity between classes during the early stages of disease development.

4.4.5 *t*-SNE visualization

In addition to classification metrics, the quality of the feature representations learned by the model was analyzed using *t*-SNE. This method projects high-dimensional embeddings into a two-dimensional space, allowing the distribution of samples across classes to be visually examined. *t*-SNE visualization was used to evaluate the backbone’s ability to generate feature representations that exhibit intra-class compactness and inter-class separation. As shown in Figure 8, the ResNet-50 backbone can form three main clusters corresponding to the Healthy, Pest, and Virus classes. In general, each class is fairly well separated. However, some samples remain outside their primary clusters, particularly in the Virus class, where samples appear near the Healthy and Pest clusters. The presence of these outliers indicates that the feature representations generated by ResNet-50 are not yet fully capable of consistently separating all samples. This suggests that some images still share similar features, which could increase the likelihood of misclassification during prediction.

To evaluate the impact of the proposed backbone, a similar analysis was conducted on the FOMAML model using ResNet-50Plus. Figure 9 shows that the embeddings generated by ResNet-50Plus exhibit a more structured distribution than those from ResNet-50. The three classes form tighter clusters with greater spacing between them, resulting in clearer inter-class separation. Furthermore, the number of samples outside the main clusters has decreased significantly. Only a few points remain close to other clusters, while the majority of samples are now concentrated within their respective groups. These results indicate that integrating ResNet-50Plus with the ECA module improves feature representation by emphasizing more relevant channel information, thereby producing more discriminative embeddings. These findings are consistent with the improved classification performance observed in the earlier experiments, in which the model achieved higher accuracy and exhibited more clearly separated feature representations across classes.

4.4.6 Grad-CAM analysis

The Grad-CAM approach was employed to analyze the visual regions on which the model focused during the classification process. Visualizations were generated using the FOMAML framework with the ResNet-50 and ResNet-50Plus backbones for each class. The Grad-CAM visualization results are presented in Figure 10. The Grad-CAM analysis in Figure 10 shows differences in the focus of activation between ResNet-50 and ResNet-50Plus. With ResNet-50, the activation areas appeared relatively dispersed, and in some samples, the activations were not fully centered on the primary symptom areas. In contrast, ResNet-50Plus produced activations that were more precisely localized to relevant leaf regions, such as areas exhibiting curling, textural changes, and discoloration associated with viral infection. This distinction is particularly evident in the Pest and Virus classes, which

share morphological similarities. This pattern is consistent with the confusion matrix results, in which the use of ResNet-50Plus reduced classification errors for both classes.

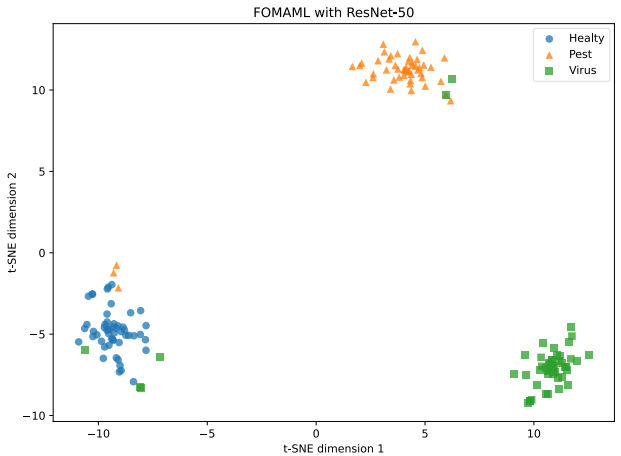


Figure 8. Visualization of embeddings using *t*-SNE on ResNet-50

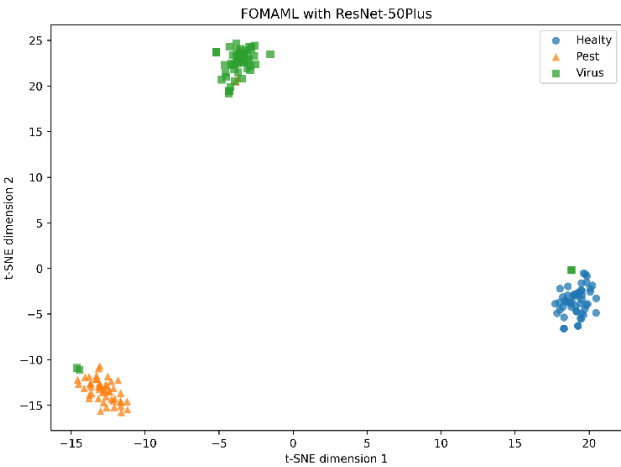


Figure 9. Visualization of embeddings using *t*-SNE on ResNet-50Plus

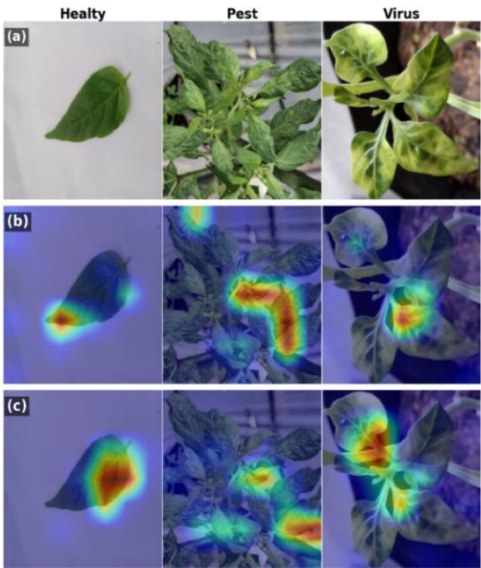


Figure 10. Gradient-weighted Class Activation Mapping (Grad-CAM) visualization: (a) original images, (b) ResNet-50, (c) ResNet-50Plus

5. CONCLUSION

This study proposes ResNet-50Plus, a residual backbone based on ResNet-50 that has been enhanced by replacing the ReLU activation function with LeakyReLU across all residual blocks and integrating ECA to support the few-shot classification of chili leaf curl disease using meta-learning. The proposed backbone is evaluated within three meta-learning frameworks, namely FOMAML, ProtoNet, and RelationNet, using identical experimental configurations.

Experimental results based on ten random seeds demonstrate that ResNet-50Plus consistently outperforms ResNet-50 across all shot settings. Within the FOMAML framework, the proposed backbone achieves its best performance in the 10-shot setting, with low performance variance across random seeds and a statistically significant improvement over the baseline. The aggregate confusion matrices further reveal a reduction in classification errors, particularly for the Pest and Virus classes. Meanwhile, Grad-CAM and *t*-SNE visualizations provide complementary qualitative evidence, revealing more focused activation patterns and embedding clusters that are denser and more clearly separated. Overall, these results indicate that ResNet-50Plus consistently improves classification performance in few-shot learning scenarios while maintaining relatively low variance across random seeds.

Future research could focus on developing adaptive attention mechanisms and multi-scale feature representations to enhance the model's ability to capture disease-related characteristics at multiple scales. Additionally, evaluating the model on larger and more diverse datasets collected under various field conditions is necessary to further assess its generalization capability.

ACKNOWLEDGMENT

The authors would like to thank the Center for Higher Education Financing and Assessment (*Pusat Pembiayaan dan Asesmen Pendidikan Tinggi*, PPAPT), Ministry of Higher Education, Science, and Technology of Indonesia (*Kemendikristek*), for awarding the 2025 Doctoral Study Completion Scholarship, which supported the publication of this article.

REFERENCES

- [1] Perdani, A.Y., Syukur, M., Wahyuni, Y., et al. (2025). Chili (*Capsicum* sp.) shelf life and genetic variability analyses in fruit cells' structure. *SABRAO Journal of Breeding and Genetics*, 57(5): 1897-1907. <https://doi.org/10.54910/sabrao2025.57.5.11>
- [2] Paradisa, Y.B., Hidayat, S.H., Saputra, A., et al. (2024). Evaluation of three different DNA extraction methods for the detection of Pepper yellow leaf curl virus (PepYLCV) by polymerase chain reaction. *IOP Conference Series: Earth and Environmental Science*, 1377(1): 012106. <https://doi.org/10.1088/1755-1315/1377/1/012106>
- [3] Trigka, M., Dritsas, E. (2025). A comprehensive survey of deep learning approaches in image processing. *Sensors*, 25(2): 531. <https://doi.org/10.3390/s25020531>
- [4] Quiñonez-Baca, L.C., Ramirez-Alonso, G., Gaxiola, F., Manzo-Martinez, A., Cornejo, R., Lopez-Flores, D.R. (2025). A comparative evaluation of meta-learning models for few-shot chest x-ray disease classification. *Diagnostics*, 15(18): 2404. <https://doi.org/10.3390/diagnostics15182404>
- [5] Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys*, 53(3): 1-34. <https://doi.org/10.1145/3386252>
- [6] Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A. (2021). Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9): 5149-5169. <https://doi.org/10.1109/tpami.2021.3079209>
- [7] Wu, J., Xiong, W., Wang, W.Y. (2019). Learning to learn and predict: A meta-learning approach for multi-label classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, pp. 4353-4363. <https://doi.org/10.18653/v1/d19-1444>
- [8] Vettoruzzo, A., Bouguelia, M.-R., Vanschoren, J., Rögnvaldsson, T., Santosh, K. (2024). Advances and challenges in meta-learning: A technical review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7): 4763-4779. <https://doi.org/10.1109/tpami.2024.3357847>
- [9] Zhang, J., Yu, W., Yang, X., Deng, F. (2019). Few-shot learning for ear recognition. In *Proceedings of the 2019 International Conference on Image, Video and Signal Processing*, Shanghai, China, pp. 50-54. <https://doi.org/10.1145/3317640.3317646>
- [10] Sarangi, P., Agarwal, R., Basu, T. (2026). ProtoRadNet: Prototypical patches of convolutional features for radiology image classification network. *Artificial Intelligence in Medicine*, 172: 103324. <https://doi.org/10.1016/j.artmed.2025.103324>
- [11] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H.S., Hospedales, T.M. (2018). Learning to compare: Relation network for few-shot learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 1199-1208. <https://doi.org/10.1109/cvpr.2018.00131>
- [12] Suwarningsih, W., Kirana, R., Khotimah, P.H., et al. (2024). Emphasizing data quality for the identification of chili varieties in the context of smart agriculture. *Interdisciplinary Journal of Information, Knowledge, and Management*, 19: 005. <https://doi.org/10.28945/5257>
- [13] Tak, J.H., Hong, B.W. (2024). Enhancing model agnostic meta-learning via gradient similarity loss. *Electronics*, 13(3): 535. <https://doi.org/10.3390/electronics13030535>
- [14] Yuan, C., Jia, X., Wang, L., Yang, C. (2025). Fine-grained prototype network for MRI sequence classification. *Current Medical Imaging*, 21: E15734056361649. <https://doi.org/10.2174/0115734056361649250717162910>
- [15] Xu, Y., Wang, J., Chen, Y., Wang, J., Sun, H., Fang, Z. (2026). ProtoMark: A transfer prototype network with Markov-enhanced features for few-shot intrusion detection. *Applied Soft Computing*, 192: 114806. <https://doi.org/10.1016/j.asoc.2026.114806>

- [16] Suwarningsih, W., Kirana, R., Khotimah, P.H., et al. (2025). Chili leaf segmentation using meta-learning for improved model accuracy. *Bulletin of Electrical Engineering and Informatics*, 14(3): 2209-2221. <https://doi.org/10.11591/eei.v14i3.7929>
- [17] Ren, J., An, Y., Lei, T., et al. (2026). Adaptive feature selection-based feature reconstruction network for few-shot learning. *Pattern Recognition*, 171: 112289. <https://doi.org/10.1016/j.patcog.2025.112289>
- [18] Wang, M., Leng, Z., Hu, P., et al. (2026). FRCAE: Feature regularization meta-learning with channel-wise attention expansion. *Pattern Recognition*, 171: 112334. <https://doi.org/10.1016/j.patcog.2025.112334>
- [19] Li, X., Zhu, G., Hu, A., Xing, L., Xiang, L. (2025). A meta-learning method based on meta-feature enhancement for bearing fault identification under few-sample conditions. *Mechanical Systems and Signal Processing*, 226: 112370. <https://doi.org/10.1016/j.ymssp.2025.112370>
- [20] Xu, J., Li, Z., Du, B., Zhang, M., Liu, J. (2020). Reluplex made more practical: Leaky ReLU. In 2020 IEEE Symposium on Computers and Communications (ISCC), Rennes, France, pp. 1-7. <https://doi.org/10.1109/iscc50000.2020.9219587>
- [21] Padshetty, S., Ambika. (2023). Leaky ReLU-ResNet for plant leaf disease detection: A deep learning approach. *Engineering Proceedings*, 59(1): 39. <https://doi.org/10.3390/engproc2023059039>
- [22] Chen, X., Cheng, Y., Wang, S., Sang, G., Nah, K., Wang, J. (2025). A periodic mapping activation function: Mathematical properties and application in convolutional neural networks. *Mathematics*, 13(17): 2843. <https://doi.org/10.3390/math13172843>
- [23] Singh, M.K., Mishra, S., Tiwari, U., Patel, V.K., Maurya, S. (2024). Analyzing the effects of alpha variation in exponential linear unit activation function for deep neural networks. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, pp. 1-6. <https://doi.org/10.1109/icccnt61001.2024.10723936>
- [24] Chieng, H.H., Wahid, N., Pauline, O., Perla, S.R.K. (2018). Flatten-T Swish: A thresholded ReLU-Swish-like activation function for deep learning. *International Journal of Advances in Intelligent Informatics*, 4(2): 76-86. <https://doi.org/10.26555/ijain.v4i2.249>
- [25] Chen, X., Lin, P. (2024). Few-shot image classification based on multimodal information fusion. In 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), Shenzhen, China, pp. 657-660. <https://doi.org/10.1109/icicml63543.2024.10957775>
- [26] Yang, C., Zhang, S., Liu, Z. (2026). Few-shot classification using ensemble of multi-scale median-enhanced features. *International Journal of Network Dynamics and Intelligence*, 5(1): 5. <https://doi.org/10.53941/ijndi.2026.100005>
- [27] Satrya, W.F., Yun, J.H. (2023). Combining model-agnostic meta-learning and transfer learning for regression. *Sensors*, 23(2): 583. <https://doi.org/10.3390/s23020583>
- [28] Ji, Z., Chai, X., Yu, Y., Pang, Y., Zhang, Z. (2020). Improved prototypical networks for few-shot learning. *Pattern Recognition Letters*, 140: 81-87. <https://doi.org/10.1016/j.patrec.2020.07.015>
- [29] Saputra, R.A., Wajhillah, R., Farlina, Y., Noviani, H., Nurusysyifa, S. (2026). Data augmentation effects on ProtoNet few-shot yellow disease severity in chili leaves. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, 11(3): 882-890. <https://doi.org/10.33480/jitk.v11i3.7458>
- [30] Wu, Y., Li, Y., Zhao, T., et al. (2023). Improved prototypical network for active few-shot learning. *Pattern Recognition Letters*, 172: 188-194. <https://doi.org/10.1016/j.patrec.2023.06.015>
- [31] Rachmad, A., Fuad, M., Rochman, E.M.S. (2023). Convolutional neural network-based classification model of corn leaf disease. *Mathematical Modelling of Engineering Problems*, 10(2): 530-536. <https://doi.org/10.18280/mmep.100220>
- [32] Anton, Rustad, S., Shidik, G.F., Syukur, A. (2026). A transfer learning-based hybrid deep learning framework for multi-crop plant leaf disease classification. *Ingénierie des Systèmes d'Information*, 31(2): 357-369. <https://doi.org/10.18280/isi.310204>
- [33] Li, R., Salih, S., Hurtt, S. (2004). Detection of Gemini viruses in sweet potato by polymerase chain reaction. *Plant Disease*, 88(12): 1347-1351. <https://doi.org/10.1094/pdis.2004.88.12.1347>
- [34] Koeda, S., Homma, K., Tanaka, Y., et al. (2018). Inoculation of Capsicums with Pepper yellow leaf curl Indonesia virus by combining agroinoculation and grafting. *The Horticulture Journal*, 87(3): 364-371. <https://doi.org/10.2503/hortj.okd-137>
- [35] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, pp. 11531-11539. <https://doi.org/10.1109/cvpr42600.2020.01155>