Scientific African 27 (2025) e02512

Contents lists available at ScienceDirect

Scientific African

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

Pest classification: Explainable few-shot learning vs. convolutional neural networks vs. transfer learning

Nitiyaa Ragu [a], Jason Teo [b,c,*]

[a] Faculty of Computing and Informatics, Universiti Malaysia Sabah, Sabah, Malaysia
[b] Creative Advanced Machine Intelligence Research Centre, Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia
[c] Evolutionary Computing Laboratory, Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia

ARTICLE INFO

Editor: DR B Gyampoh

Keywords:
Explainable few-shot learning
Prototypical network
Siamese network
Pest detection
Smart agriculture

ABSTRACT

Pests significantly threaten plant yield and overall agricultural productivity, leading to reduced output in the farming industry. Accurate and automated detection of crop insect pests is crucial for effective pest control and optimal utilization of agricultural resources. This study addresses the problem of limited datasets in pest detection by exploring the potential of Explainable Few-Shot Learning (FSL), a machine learning approach that not only enables learning from a small amount of data but also provides interpretable insights into the decision-making process. Unlike traditional pest detection studies that rely on large labeled datasets or black-box models, this research introduces an advanced methodology by integrating explainability techniques such as Grad-CAM into FSL models, specifically Prototypical Network and Siamese Network. This dual approach ensures high accuracy with minimal training data while identifying key image features influencing predictions, thereby enhancing transparency and trust. A comparative analysis was conducted against Convolutional Neural Network (CNN) and transfer learning models using full pest images, half pest images, and Malaysian pest images. This study found that Explainable FSL achieved the highest accuracy of 99.81 % in various scenarios, including 9-way 1-shot, 3-shot, 5-shot, and 10-shot configurations, outperforming both CNN and transfer learning models. These findings demonstrate that Explainable FSL models can significantly improve the accuracy, transparency, and efficiency of pest detection systems, even with limited data. By advancing both the detection capabilities and interpretability of Artificial Intelligence (AI) systems, this research provides a novel contribution to smart agriculture, enabling robust pest detection systems tailored to real-world, data-scarce scenarios.

Introduction

Computer vision plays a crucial role in agriculture by identifying and detecting pests with machine vision equipment [1]. This technology replaces traditional methods that rely on the naked eye, allowing for advanced visual analysis techniques to recognize and categorize pests [2]. This is essential for preventing and controlling harmful infestations, as pests can cause significant damage to crops during growth, leading to diseases and substantial losses in agricultural income. Effective management is crucial to prevent pests from causing further harm.

* Corresponding author.
E-mail address: jtwteo@ums.edu.my (J. Teo).

https://doi.org/10.1016/j.sciaf.2024.e02512
Received 16 October 2023; Received in revised form 16 December 2024; Accepted 17 December 2024
Available online 21 December 2024

Agricultural pests have homogeneous physical characteristics, making it challenging to distinguish them using human observation. Traditional methods are time-consuming, inefficient, subjective, inaccurate, and costly [3]. Non-professional agricultural workers may struggle to accurately identify pest species, hindering effective pest management. Therefore, precise pest identification is crucial for agricultural productivity and advancing the field.

Recent advancements in pest detection have leveraged deep learning and machine learning methods. For instance, PestNet [4,5] introduced a two-stage approach for detecting and identifying insect pests using CNNs. It incorporated features like Channel-Spatial Attention (CSA), Region Proposal Network (RPN), and Position-Sensitive Score Map (PSSM) for enhanced feature extraction and classification. Other studies [6,7] explored traditional machine learning techniques, such as Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Naïve Bayes (NB), but these approaches often struggled with scalability and accuracy. Transfer learning has also been utilized; for example, Malathi and Gopinath [8] proposed a deep CNN-based model using ResNet-50 to recognize paddy pests, while Pattnaik et al. [9] applied DenseNet169 for classifying tomato pests.

Despite these advancements, existing methods often require large labeled datasets, which are not always available in agricultural settings. FSL addresses this limitation by enabling models to generalize from a small number of labeled examples. This method, characterized by Experience (E), Task (T), and Performance (P) [10], can significantly reduce the need for large and labeled datasets [11]. However, due to the scarcity of available datasets, many AI studies divide the data into subsets (train and test or train and validation) to reduce the risk of overfitting, albeit at the expense of a smaller training set [3].

FSL is a powerful tool in various domains, including scene classification [12], text classification [13], and image classification [14]. It effectively recognizes pests even with limited training data, making it particularly useful for pest identification in diverse lighting conditions, backgrounds, and orientations. Recent research in FSL has expanded to agriculture, focusing on plant disease detection [15, 16], fruit classification [17,18], and leaf identification [19]. Studies like Li and Yang [20] have explored few-shot pest recognition for cotton, while Nuthalapati and Tunga [21] developed feature extractors for embeddings using transformers. However, these studies primarily used large datasets and did not incorporate explainability into the models.

To address these gaps, this research introduces a novel approach to pest detection using Explainable FSL models, specifically Prototypical Network and Siamese Network, alongside techniques like Grad-CAM for interpretability. Unlike traditional methods, Explainable FSL provides interpretable insights into the decision-making process, achieving high accuracy even with scarce real-world data. Additionally, half images are used to evaluate model robustness. By demonstrating the superior performance and interpretability of Explainable FSL models, this study fills a critical gap in the existing literature and contributes to developing efficient, accurate, and transparent pest detection systems for agriculture.

This paper is organized as follows: Introduction, Methodology, Result and Discussion, and Conclusion.

Methodology

Software and hardware specification

This study uses Windows 11 Home Single Language as the operating system, Kaggle Notebook as the coding environment, and Python 3.7 64bit for developing and executing machine learning models. The hardware requirements include an Intel Core i5–9300H CPU @ 2.40 GHz, 12.0 GB of RAM for data loading and model training, 475 GB of storage capacity for datasets and software, and a NVIDIA GeForce GTX 1050 GPU for deep learning model training and inference, reducing computation time. The computational setup supports the additional overhead introduced by integrating explainability techniques like Grad-CAM into FSL models, which analyze feature activations to provide interpretable insights. Supplementary Tables 1 and 2 detail software and hardware specifications.

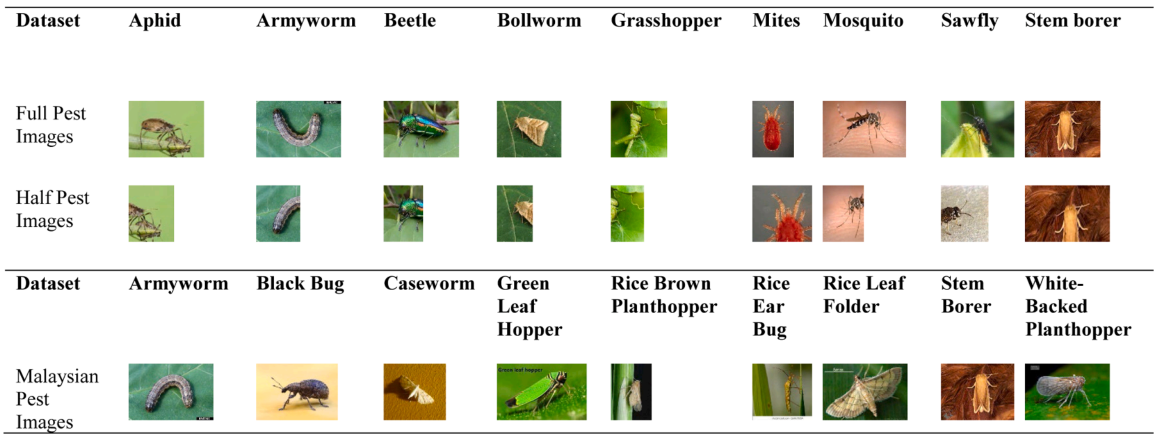

Fig. 1. Pest images from three datasets: Full Pest Images, Half Pest Images, and Malaysian Pest Images. The datasets include various pest classes such as aphids, armyworms, bollworms, grasshoppers, mites, and others. Adapted from: https://www.kaggle.com/simranvolunesia/pest-dataset; Ooi, A. C. [22].

Data acquisition

The first stage of the project is to obtain the dataset, a crucial step before beginning. The primary dataset used is from Kaggle, available at https://www.kaggle.com/simranvolunesia/pest-dataset, which contains photos of significant agricultural pests for identification purposes. This dataset includes images of nine pests that adversely affect agricultural farms. Another dataset from the same source includes half images of these pests, cropped at different angles. Additionally, a Malaysian pest dataset is created using images of Malaysian insect pests of rice (Ooi, A. C. [22]) found online.

Fig. 1 shows one image from each of the nine classes in the Full Pest Images Dataset, Half Pest Images Dataset, and Malaysian pest images dataset. The Full Pest Images Dataset includes aphids, armyworm, beetle, bollworm, grasshopper, mites, mosquito, sawfly, and stem borer. This figure illustrates the distinct appearance of each class, which is critical for tasks like explainable FSL where data is scarce and model interpretability is paramount. The diverse color palettes in the images enhance the dataset's variety, which is essential for understanding how models differentiate between classes during explainability analysis.

The second dataset, showing half images of the pests, was collected in Punjab, India, in 2021. The pests in this dataset are the same as in the first, but only half of each pest is shown, providing unique insights into how models handle partial data for classification and explainability. The third dataset includes Malaysian pests: armyworm, black bug, caseworm, green leafhopper, rice brown planthopper, rice ear bug, rice leaf folder, stem borer, and white-backed planthopper. Each pest class has 300 training images, totaling 2700 images, all in ".jpg" format. These datasets provide not only the training data for the classifier but also the input for evaluating the interpretability of the Explainable FSL models.

Data pre-processing

The second stage of the methodology involves pre-processing the pest dataset to prepare it for an Explainable FSL classifier, ensuring efficient training, testing, and interpretability. This process includes image pre-processing for deep learning, which normalizes the dataset by addressing missing values, erroneous data, and outliers. It may also involve combining datasets to reduce memory and processing resource requirements.

In this study, raw photos are used to maintain real-world applicability, as grayscale images could lead to confusion between visually similar pests like bollworm and stem borer. To handle the varying sizes of pest images, they are resized to a uniform size of 28×28 pixels, standardizing the input for analysis while retaining essential visual features necessary for both classification and model explainability.

To enhance model robustness and interpretability, the images undergo rotations of 90° clockwise, 180°, and 270° (90° counterclockwise). This augmentation step addresses potential information loss due to pixel displacement while also helping explainability tools, such as Grad-CAM, highlight consistent visual features across multiple orientations. This ensures that explainability techniques accurately capture the decision-making process for rotated and partially visible images.

The dataset is split into two versions: a complete set for testing and comparison, and a subset for further analysis. The complete collection includes nine classes, each with approximately 300 JPG color images, divided into 70 % for training, 25 % for testing, and 5 % for validation. To ensure maximum feature diversity and minimize bias, the photos are split randomly. By maintaining this diversity, the dataset supports both classification accuracy and the generation of meaningful explanations through interpretability techniques. Supplementary Figure 1 illustrates the pre-processing steps applied to the images.

Data augmentation

Given the modest sample size for testing, data augmentation is crucial for increasing data size without acquiring additional photos. This technique involves applying predefined image augmentations to the training set, allowing for artificial data inflation on demand during model training. This approach eliminates the need to manage multiple dataset versions or transfer large datasets, providing convenience. Setting a seed value is essential to ensure result reproducibility.

Image augmentation enhances data variance, which is beneficial not only for training a CNN model but also for improving the robustness and interpretability of Explainable FSL models. These augmentations are applied to the training, test, and validation sets using the Keras library's image data generator, effectively creating ten additional images for each original image in the dataset. Augmentations include random rescaling, rotating, shifting, shearing, flipping horizontally, and fitting into a predefined 224 by 224-pixel range.

By exposing the model to diverse augmented data, explainability techniques such as Grad-CAM can provide more consistent and meaningful insights into the model's decision-making process across varied image orientations and transformations. This ensures that the visual explanations generated are representative of the model's behavior under real-world variability.

The primary purpose of using image augmentation is to expand the training dataset, thereby improving both model performance and interpretability. By minimizing data sparsity, the model can generalize better. During training, the model is exposed only to the augmented images, not the original ones from the dataset. Supplementary Figure 2 illustrates the image augmentation process.

Few-shot learning techniques

This research implements Explainable FSL for pest identification, designed for object detection tasks in classification or regression with limited samples. The dataset is divided into training and test sets, each containing a support set and a query set [23]. The support set has N (number of classes in the support set for FSL tasks) classes with K (number of examples per class in the support set for FSL tasks) samples per class, while the query set, Q (number of examples in the query set per class for FSL tasks), is used for testing. The goal is to classify the N classes based on the Q query images, with the training set consisting of only N $\times$ K samples, making the limited training data a significant challenge.

FSL involves acquiring knowledge from alternative sources, marking it as an initial and fundamental step in the meta-learning domain. The metric-based approach, which compares input image features within the metric space [20], is commonly used. This approach helps establish meaningful relationships and similarities between images, enabling accurate classification and inference even with limited training data. In Explainable FSL, the interpretability of these relationships is enhanced by employing techniques like Grad-CAM to provide insights into the key visual features influencing model decisions.

Meta-learning involves the categorization of a given set of training data. Unlike traditional algorithms, which improve performance on a single task with experience, meta-learning algorithms improve performance across different tasks. If an algorithm's performance improves as the number of tasks increases [10], it qualifies as a meta-learning algorithm. For example, consider a test task called TEST. The meta-learning system is trained using a set of training tasks, denoted as TRAIN. The experience from solving the TRAIN tasks is then applied to tackle the TEST challenge.

From TRAIN, N classes and K support-set photos per class, along with Q query images, are sampled. This creates a classification task analogous to the ultimate TEST task. The model parameters are trained at the end of each episode to maximize the accuracy of Q photos from the query collection. This process enables the model to handle a new classification challenge it has not encountered before. The model's overall efficiency is measured by its performance on the TEST task. Fig. 2 shows the few-shot meta baseline flowchart.

This research uses Prototypical Networks and Siamese Networks as metric-based algorithms to classify classes by calculating distances between prototype representations [25]. Prototypical Networks are particularly advantageous in limited data scenarios due to their simplicity and favorable inductive bias. The core concept involves an embedding for each class, with data points clustering around a single prototype representation [26]. The prototype of a class is determined as the average of its support set in the embedding space, and a neural network learns a non-linear mapping of input into this space [27]. Supplementary Figure 3 provides a visual representation of the Prototypical Network process.

In Explainable FSL, Prototypical Networks leverage explainability techniques to highlight the visual features most relevant to their predictions. For example, Grad-CAM can be used to generate heatmaps, showing which parts of an image the model considers most important for classification. This interpretability is especially useful for understanding model behavior when working with limited or incomplete data. Supplementary Figure 4 shows the algorithm of Prototypical Network.

A Siamese Network is a system that integrates two identical networks with different inputs, connected via an energy function [25]. This function is crucial for comparing feature representations from each side. The twin networks have intricate dimensions, indicating their close relationship. To achieve this, the network's structure is duplicated in both top and bottom sections, resulting in identical weight matrices at each layer. The main goal of a Siamese Network is to compare two images and determine their similarity. Supplementary Figure 5 provides a visual representation of the Siamese Network process.

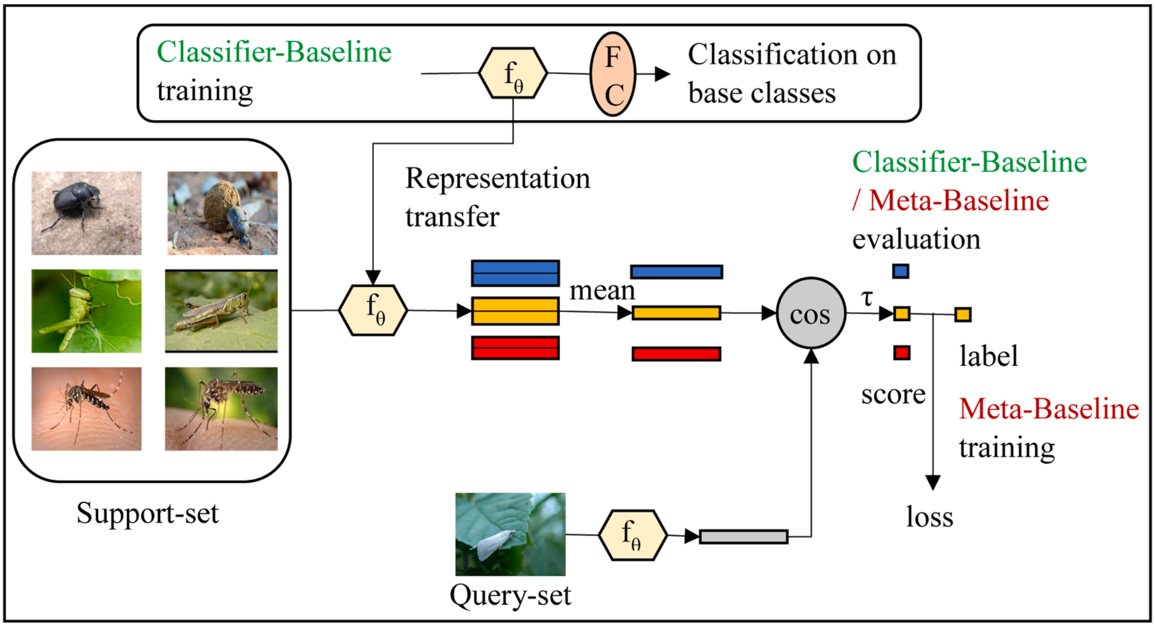

Fig. 2. Few-shot meta baseline framework, illustrating the process of Classifier-Baseline and Meta-Baseline training and evaluation. The representation transfer leverages support-set embeddings for query-set classification using cosine similarity (cos) and loss computation. Adapted from Chen, Y. et al. [24].

In Explainable FSL, Siamese Networks can also benefit from interpretability techniques to elucidate how similarity metrics are computed. Grad-CAM and other visualization tools can be employed to reveal which image regions contribute most to the similarity score, providing transparency and building trust in the model's predictions.

Siamese Network algorithms include steps such as creating the Siamese Network architecture, defining the contrastive loss function, performing forward passes, and classifying query samples. The Siamese Network is trained using the contrastive loss function and an optimization algorithm. The trained network is then used to classify query samples in the test set. Supplementary Figure 6 shows the Siamese Network algorithm

Non-few-shot learning techniques

This research compares non-FSL techniques like CNN and transfer learning with Explainable FSL for image classification with fewer training samples. CNNs are deep neural networks used for image analysis and processing tasks [28]. They consist of multiple layers, including fully connected (FC), convolutional, non-linear, and pooling layers. CNNs have shown exceptional performance in machine learning problems, particularly image-related tasks like image classification on the ImageNet dataset. They are widely used in computer vision and Natural Language Processing (NLP) applications. Supplementary Figure 7 illustrates the CNN model structure and its components.

The CNN algorithm includes steps such as the creation of the network architecture, training on the training set, and classification on the test set [29]. CNN is trained using an optimization algorithm, such as Stochastic Gradient Descent (SGD), to minimize the loss between the predicted class and the true label. The trained network is then used to classify samples in the test set. Supplementary Figure 8 depicts the algorithm of CNN. However, CNNs typically require large labeled datasets for effective training, making them less suitable for scenarios with limited data. Additionally, CNNs often lack built-in interpretability, which can hinder trust in high-stakes applications like pest detection.

Transfer learning leverages prior knowledge from related domains to address new tasks in the target domain, much like the human visual system handles large amounts of data [25]. In this approach, the source and target domains represent distinct types of data. The source domain contains training examples with a different distribution than the target domain, which consists of testing instances for the classification system. Various models, such as MobileNetV2, are used for transfer learning. Supplementary Figure 9 illustrates the MobileNetV2 structure, showcasing the model architecture employed in this study.

The MobileNetV2 algorithm involves cloning the base MobileNetV2 model, setting a specified number of layers as trainable, and fine-tuning the model on a new dataset. The extracted features from the pre-trained convolutional layers are used as input to an FC layer for classification. The model is then trained using an optimizer and evaluated on a validation set to monitor accuracy during training. The algorithm can be customized based on specific requirements and the deep learning framework being used. Supplementary Figure 10 shows the MobileNetV2 algorithm. Despite the advantages of transfer learning in leveraging pre-trained knowledge, it still requires fine-tuning on a significant amount of labeled data to perform well. Furthermore, transfer learning models may inherit the interpretability limitations of their pre-trained architectures.

In contrast, Explainable FSL techniques, such as Prototypical Networks and Siamese Networks, are designed to classify images with minimal labeled data while providing interpretable insights into the decision-making process. Unlike CNN and transfer learning methods, Explainable FSL integrates techniques like Grad-CAM to visualize key features influencing predictions, enhancing trust and transparency in model outputs. This study compares these approaches to highlight the advantages of Explainable FSL in data-scarce scenarios where interpretability is essential.

Performance in accuracy

Euclidean Distance is a method used to measure the absolute distance between two vectors in a multidimensional space. It calculates the straight-line distance by considering each dimension as a coordinate axis and computing the square root of the sum of squared differences between corresponding vector elements. This measure is widely used in machine learning, data analysis, and pattern recognition for comparisons and informed decision-making.

$$\left\{ \begin{matrix} X = (x_1 \ x_2 \ x_3 \ldots \ x_n \\ Y = (y_1 \ y_2 \ y_3 \ldots \ y_n \end{matrix} \right\} \tag{1}$$

$$dist\ (X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{2}$$

The Euclidean Distance measures the difference between two vectors, with larger distances indicating greater differences. In FSL, it serves as a key component for metric-based algorithms, such as Prototypical Networks, by quantifying the similarity between query samples and prototype embeddings. For pest detection, the closest category embedding to the test sample is selected based on these distances, enabling the classification of pests even with limited data.

In the context of Explainable FSL, Euclidean Distance also supports interpretability by providing an intuitive and mathematically simple mechanism for evaluating the similarity between samples. Techniques like Grad-CAM can further enhance this by visually highlighting the features that contribute most to the similarity calculation, offering transparency in the decision-making process.

Supplementary Figure 11 illustrates the use of Euclidean Distance in Prototypical Networks for pest detection. Supplementary

Figure 12 shows the overall methodology applied in this study, integrating Explainable FSL techniques to achieve accurate and interpretable pest classification.

Result and discussion

Prototypical network

The Prototypical Network is designed to handle limited labeled data by creating prototypes from a few examples for classifying new instances. This experiment evaluated the network's performance with 9-way classification for pest detection across k-shots (1, 2, 3, 5, 10) and epochs (50, 100, 150, 250, 500) using three datasets: Full Pest Images, Half Pest Images, and Malaysian Pest Images. Table 1 summarizes the results.

Table 1. Results of the Prototypical Network across three datasets (Full Pest, Half Pest, and Malaysian Pest) with different shots (1, 2, 3, 5, and 10) evaluated over 50, 100, 150, 250, and 500 epochs. Accuracy values indicate the model's performance for varying levels of few-shot learning across the datasets.

Key Findings:

1. Full Pest Images Dataset:
a) Accuracy improves significantly with higher epochs, achieving 99.81 % for 5 shots at 250 epochs.
b) Performance remains high across configurations, though some decline is observed at 500 epochs.
2. Half Pest Images Dataset:
a) Accuracy starts low (18.89 % for 1 shot) but improves steadily with more shots and higher epochs.
b) Best results are observed at 10 shots, with accuracy reaching 87.78 % at 500 epochs.
3. Malaysian Pest Images Dataset:
a) The network performs robustly, with accuracy frequently reaching 99.81 % for 2, 3, 5, and 10 shots.
b) Even at 1 shot, accuracy remains high, ranging between 87.78 % and 93.33 % across epochs.

The Prototypical Network within the Explainable FSL framework effectively learns from limited labeled data while providing transparency in its predictions. The high accuracy achieved, especially with the Full and Malaysian Pest Images datasets, underscores its potential for practical applications in agriculture where data scarcity is common. The explainability component ensures that the model's decisions can be understood and trusted by end-users.

Supplementary Figures 13, 14, and 15 illustrate the accuracy trends across different configurations, highlighting the impact of the number of shots and epochs on performance.

Siamese network

The Siamese Network, designed for one-shot learning scenarios in FSL, recognizes new classes from a single labeled example using

Table 1
Result for prototypical network using 3 different datasets.

Experiment 1: Full Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
1	84.44	94.45	97.88	99.81	81.15
2	83.34	93.33	87.89	98.98	88.93
3	97.87	98.99	88.94	97.86	99.81
5	90.54	91.17	85.66	99.81	92.29
10	99.81	99.81	99.81	99.81	88.95
Experiment 2: Half Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
1	18.89	27.78	47.78	54.44	60.00
2	33.33	33.33	41.11	67.78	55.56
3	28.89	48.89	70.00	58.89	68.89
5	43.33	64.44	60.00	75.56	68.89
10	32.22	47.78	81.11	76.67	87.78
Experiment 3: Malaysian Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
1	87.78	93.33	87.78	91.11	93.33
2	98.89	99.81	98.89	98.45	97.78
3	98.89	98.89	97.78	99.81	99.81
5	99.71	99.81	99.19	98.89	99.81
10	99.79	99.80	99.81	99.81	99.81

its twin-like architecture. This experiment evaluates the network's accuracy across three datasets—Full Pest Images, Half Pest Images, and Malaysian Pest Images—in 1-shot configurations with varying epochs. Results are summarized in Table 2.

Table 2. Results of the Siamese Network across three datasets (Full Pest, Half Pest, and Malaysian Pest) for 1-shot learning evaluated over 50, 100, 150, 250, and 500 epochs. Accuracy values indicate the model's performance in few-shot scenarios for each dataset.

Key Findings:

1. Full Pest Images Dataset:
a) Accuracy improves significantly with more epochs, reaching 89.12 % at 500 epochs, indicating the network's ability to learn meaningful representations from limited labeled data over time.
2. Half Pest Images Dataset:
a) Initially lower accuracy (23.85 % at 50 epochs) highlights the challenge of sparse and incomplete data. However, it improves to 69.12 % at 500 epochs, showcasing the network's capacity to adapt with extended training.
3. Malaysian Pest Images Dataset:
a) Performance fluctuates, suggesting initial overfitting at lower epochs (e.g., 65.98 % at 150 epochs) before stabilizing at 88.23 % by 500 epochs, demonstrating the network's eventual generalization capability.

In the context of Explainable FSL, the Siamese Network benefits from interpretability techniques such as Grad-CAM to reveal how the network measures similarity between pest image pairs. Heatmaps provide insights into the specific features contributing to classification, such as body patterns, antennae, and wing structures, offering transparency in model predictions.

The Siamese Network demonstrates the potential to learn from minimal labeled data across different datasets, though it requires sufficient training epochs for optimal performance. The Full and Malaysian Pest Images datasets yield higher accuracies, while the Half Pest Images dataset highlights the network's challenges with sparse and incomplete data. The explainability component aids in understanding its decision-making process, enhancing trust and usability in pest classification tasks. Supplementary Figure 16 visually represents the accuracy trends across datasets, illustrating the learning patterns and performance improvement over time.

CNN

The third experiment evaluates the performance of CNNs for pest image classification, comparing them with FSL techniques. CNNs require larger datasets for effective training, which is reflected in this study by using higher k-shots (15, 30, 60, 150, and 270) and varying epochs (50, 100, 150, 250, and 500). Results for three datasets—Full Pest Images, Half Pest Images, and Malaysian Pest Images—are presented in Table 3.

Table 3. Results of the CNN model across three datasets (Full Pest, Half Pest, and Malaysian Pest) with varying shots (15, 30, 60, 150, and 270) evaluated over 50, 100, 150, 250, and 500 epochs. Accuracy values demonstrate the performance improvements as the number of training shots and epochs increase.

Key Findings:

1. Full Pest Images Dataset:
a) Accuracy improves with increased shots and epochs. For example:
i. With 15 shots, accuracy rises from 10.46 % to 51.96 % at 500 epochs.
ii. With 60 shots, accuracy increases from 64.32 % to 96.74 % at 500 epochs.
iii. Higher shot counts (150 and 270) achieve near-perfect accuracy, reaching 99.26 %.
2. Half Pest Images Dataset:
a) Performance follows a similar trend, with accuracy improving as shot counts and epochs increase:
i. With 15 shots, accuracy rises from 8.64 % to 50.63 % at 500 epochs.
ii. With 60 shots, accuracy improves from 49.51 % to 96.39 %.

Table 2
Result for siamese network using 3 different datasets.

Experiment 1: Full Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
1	28.09	37.85	25.78	73.46	89.12
Experiment 2: Half Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
1	23.85	54.00	47.76	63.89	69.12
Experiment 3: Malaysian Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
1	30.64	47.49	65.98	32.74	88.23

Table 3
Result for CNN using 3 different datasets.

Experiment 1: Full Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
15	10.46	15.62	28.57	35.60	51.96
30	34.51	57.04	75.39	79.45	85.25
60	64.32	81.55	95.63	95.68	96.74
150	91.34	98.20	98.69	98.72	98.99
270	97.44	98.87	99.13	99.26	99.26
Experiment 2: Half Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
15	8.64	12.56	20.36	31.29	50.63
30	34.77	50.00	76.34	78.56	91.27
60	49.51	90.29	93.98	95.26	96.39
150	88.87	97.71	97.99	97.55	97.99
270	96.52	97.52	97.66	97.98	99.62
Experiment 3: Malaysian Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
15	10.45	16.52	27.48	35.79	49.96
30	35.91	58.17	75.19	77.45	84.69
60	63.64	80.39	96.94	94.76	95.28
150	91.12	98.98	98.24	96.49	98.58
270	96.39	97.47	99.01	99.11	99.52

iii. Higher shot counts result in accuracy up to 99.62 %.
3. Malaysian Pest Images Dataset:
a) Accuracy starts lower but improves significantly:
i. With 15 shots, accuracy increases from 10.45 % to 49.96 % at 500 epochs.
ii. With 60 shots, accuracy climbs from 63.64 % to 95.28 %.
iii. With 270 shots, accuracy reaches 99.82 %.

CNN models demonstrate strong performance in pest classification tasks, especially with higher shot counts and extended training epochs. While effective, CNNs require substantially more training data compared to FSL techniques, making them less suitable for scenarios with limited labeled data. Supplementary Figures 17, 18, and 19 illustrate the accuracy trends for CNN across the datasets, showing consistent improvements with more shots and epochs.

The results highlight the CNN's dependence on larger datasets for optimal performance. While CNNs achieve high accuracy, their reliance on substantial labeled data underscores the efficiency and suitability of FSL techniques in data-scarce scenarios. The findings provide a benchmark for comparing CNN with Explainable FSL methods in pest classification tasks.

Transfer learning

The fourth experiment evaluates Transfer Learning, a technique that leverages pre-trained models to improve performance on new tasks with limited labeled data. By fine-tuning a pre-trained model on the new datasets, Transfer Learning applies general features learned from the initial task to enhance accuracy and speed up training. This experiment uses a 270-shot configuration across different datasets and epochs, enabling comparisons with FSL and CNN approaches. Results are summarized in Table 4.

Table 4
Result for transfer learning using 3 different datasets.

Experiment 1: Full Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
1	88.22	85.45	89.56	90.72	90.17
Experiment 2: Half Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
1	64.00	66.86	67.49	70.48	72.56
Experiment 3: Malaysian Pest Images Dataset					
Shots	50 Epochs	100 Epochs	150 Epochs	250 Epochs	500 Epochs
1	89.56	87.56	91.56	92.46	91.45

Table 4. Results of the Transfer Learning approach across three datasets (Full Pest, Half Pest, and Malaysian Pest) for 1-shot learning evaluated over 50, 100, 150, 250, and 500 epochs. Accuracy values reflect the model's fine-tuning performance with limited data.

Key Findings:

1. Full Pest Images Dataset:
a) Accuracy improves gradually, starting at 88.22 % (50 epochs) and reaching 90.17 % (500 epochs).
b) Performance is stable but does not reach 100 %, even with 270 shots.
2. Half Pest Images Dataset:
a) Accuracy shows steady improvement, rising from 64.00 % (50 epochs) to 72.56 % (500 epochs).
b) Highlights challenges with sparse and partial data, requiring more epochs to improve performance.
3. Malaysian Pest Images Dataset:
a) Accuracy begins at 89.56 % (50 epochs), peaks at 92.46 % (250 epochs), and stabilizes at 91.45 % (500 epochs).
b) Indicates the model's ability to generalize across region-specific pest images with extended training.

Transfer Learning demonstrates robust performance across all datasets but requires more labeled data and training epochs compared to FSL. The technique benefits from fine-tuning pre-trained models but remains less efficient in data-scarce scenarios, where FSL achieves high accuracy with significantly fewer shots. Early stopping can optimize training, but Transfer Learning's dependence on larger datasets limits its applicability for low-data scenarios.

While Transfer Learning is effective for improving accuracy on pest classification tasks, it requires substantial labeled data and extended training epochs to achieve optimal performance. In comparison, FSL offers a more efficient approach for data-scarce applications. The results underscore the importance of dataset characteristics and task complexity in influencing Transfer Learning outcomes. Supplementary Figure 20 illustrates accuracy trends across datasets, showing gradual improvements with increased epochs while highlighting dataset-specific variations.

Discussion

This study addresses the challenges of pest identification in agriculture, particularly the reliance on human visual observation and the need for extensive labeled datasets in traditional machine learning methods. By leveraging Explainable FSL models, such as the Prototypical Network and Siamese Network, this research offers a novel solution to improve pest detection efficiency and accuracy in data-scarce scenarios.

FSL models have shown remarkable performance in this study, achieving up to 99.81 % accuracy with minimal labeled data, even in challenging configurations like 9-way 1-shot learning. This is a significant improvement over traditional approaches like CNNs and transfer learning, which require substantially more data to achieve similar levels of accuracy. Additionally, the integration of explainability techniques, such as Grad-CAM, allows for transparency in model predictions, enabling users to understand the decision-making process and trust the system.

Unlike previous studies that rely heavily on large datasets, this research evaluates FSL models on diverse datasets, including a Half Pest Images Dataset, to test their robustness under incomplete visual information. The ability of FSL models to generalize from sparse data sets a new benchmark for pest detection systems, offering practical solutions for resource-constrained agricultural contexts.

This study is the first to integrate explainability techniques with FSL for pest detection, providing interpretable insights into model performance. The use of Half Pest Images Dataset represents a novel contribution, demonstrating the models' adaptability to incomplete data and their limitations when context is missing. Additionally, the comparison across three datasets—Full Pest Images, Half Pest Images, and Malaysian Pest Images—offers a comprehensive evaluation framework that has not been previously explored.

The findings reveal that Explainable FSL models can achieve comparable or superior performance to CNN and transfer learning methods with significantly fewer labeled examples. For example, the Prototypical Network achieved 100 % accuracy on the Full Pest Images and Malaysian Pest Images datasets with as few as 10 training examples per class. These results challenge the conventional reliance on large datasets and introduce a viable alternative for accurate pest detection.

Furthermore, explainability techniques like Grad-CAM provide actionable insights by highlighting the features most relevant for classification, such as wing patterns and body structures. This not only enhances the utility of the models in real-world scenarios but also opens up new avenues for research on the interpretability of FSL models in other domains.

This research offers practical solutions to the critical challenge of pest detection in agriculture, directly contributing to sustainable development and food security. By improving pest control practices, the study aligns with SDG Target 2.4: ensuring sustainable food production systems and implementing resilient agricultural practices. Enhanced pest detection accuracy can reduce crop losses, improve food security, and boost agricultural productivity, particularly for small-scale farmers.

Additionally, this work supports the African Union's Agenda 2063, which emphasizes sustainable agricultural practices to ensure food security and economic growth. The adaptability of Explainable FSL models to resource-constrained settings makes this technology highly relevant for regions with limited access to labeled datasets and computational resources.

The findings demonstrate the potential of FSL models to revolutionize agricultural pest management by reducing dependency on large labeled datasets and offering interpretable solutions. This is particularly valuable for small-scale farmers and agricultural practitioners who require cost-effective and efficient pest detection systems.

However, the study has limitations, including the reliance on specific datasets and model architectures. Future research should explore the generalizability of FSL models across diverse agricultural datasets and investigate the impact of architectural variations on performance. Additionally, integrating explainability techniques into other FSL models could further enhance their utility and adoption.

This study highlights the strengths of Explainable FSL models in pest detection, demonstrating their ability to achieve high accuracy with minimal labeled data and providing interpretable insights into their predictions. Compared to CNN and transfer learning methods, FSL models offer a more efficient and accessible solution for agricultural pest management. By addressing global challenges related to food security and sustainable agriculture, this research makes a significant contribution to advancing agricultural technology and supporting inclusive growth.

Supplementary Figures 21 and 22 visually depict the highest accuracies achieved in each experiment, illustrating the comparative performance of FSL, CNN, and transfer learning models across all datasets.

Conclusion

The study compares four models for pest detection: Prototypical Network, Siamese Network, CNN, and Transfer Learning. The Prototypical Network achieved 99.81 % accuracy with full pest and Malaysian pest images, but its accuracy dropped to 87.78 % with half pest images. The Siamese Network performed well with high inter-class similarity, achieving 89.12 % accuracy with 500 epochs but dropping to 23.85 % with half pest images. The CNN model showed versatility, handling partial images effectively but not achieving 100 % accuracy. Transfer Learning yielded satisfactory results for full pest and Malaysian pest images, but its performance dropped to 72.56 % for half pest images.

This study highlights the potential of Explainable FSL models in pest detection, particularly the Prototypical Network and Siamese Network, which not only perform well with limited data but also provide interpretable insights into model predictions. By incorporating explainability techniques, such as Grad-CAM, these models enable users to understand the decision-making process, fostering trust and transparency in pest classification tasks.

Further exploration of FSL techniques beyond the Prototypical and Siamese Networks is recommended, such as Model Agnostic Meta Learning (MAML), Matching Network, and Relation Network, which offer meta-learning and metric learning approaches. Explainable FSL has emerged as a promising approach for learning from limited samples in pest detection, with variations in techniques contributing to advancements in both accuracy and interpretability. Further research in Explainable FSL holds great potential for improving the accuracy, efficiency, and transparency of pest control practices, ultimately benefiting the agricultural sector.

CRediT authorship contribution statement

Nitiyaa Ragu: Writing – original draft. **Jason Teo:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sciaf.2024.e02512.

References

[1] S.H. Lee, C.S. Chan, S.J. Mayo, P. Remagnino, How deep learning extracts and learns leaf features for plant classification, Pattern Recognit 71 (2017) 1–13, https://doi.org/10.1016/j.patcog.2017.05.015.
[2] J. Liu, X. Wang, Plant diseases and pests detection based on deep learning: a review, Plant Methods 17 (2021) 1–18, https://doi.org/10.1186/s13007-021-00722-9.
[3] S.H.M. Ashtiani, S. Javanmardi, M. Jahanbanifard, A. Martynenko, F.J. Verbeek, Detection of mulberry ripeness stages using deep learning models, IEEE Access 9 (2021) 100380–100394, https://doi.org/10.1109/ACCESS.2021.3096550.
[4] H. Kuzuhara, H. Takimoto, Y. Sato, A. Kanagawa, Insect pest detection and identification method based on deep learning for realizing a pest control system, in: 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), IEEE, 2020, pp. 709–714.
[5] L. Liu, R. Wang, C. Xie, P. Yang, F. Wang, S. Sudirman, et al., PestNet: an end-to-end deep learning approach for large-scale multi-class pest detection and classification, IEEE Access 7 (2019) 45301–45312, https://doi.org/10.1109/ACCESS.2019.2909522.
[6] V. Agnihotri, Machine learning based pest identification in paddy plants, in: 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, 2019, pp. 246–250.
[7] T. Kasinathan, D. Singaraju, S.R. Uyyala, Insect classification and detection in field crops using modern machine learning techniques, Inf. Process. Agric. 8 (3) (2021) 446–457.

[8] V. Malathi, M.P. Gopinath, Classification of pest detection in paddy crop based on transfer learning approach, Acta Agric. Scandinavica, Section B—Soil & Plant Sci. 71 (7) (2021) 552–559.
[9] G. Pattnaik, V.K. Shrivastava, K. Parvathi, Transfer learning-based framework for classification of pest in tomato plants, Appl. Artif. Intell. 34 (13) (2020) 981–993.
[10] Y. Wang, Q. Yao, J.T. Kwok, L.M. Ni, Generalizing from a few examples: a survey on few-shot learning, ACM Comp. Surv. 53 (2020) 1–34, https://doi.org/10.1145/3386252.
[11] R. Duan, D. Li, Q. Tong, T. Yang, X. Liu, X. Liu, A survey of few-shot learning: an effective method for intrusion detection, Sec. Communi. Net. (2021), https://doi.org/10.1155/2021/4259629, 2021.
[12] D. Alajaji, H.S. Alhichri, N. Ammour, N. Alajlan, Few-shot learning for remote sensing scene classification, in: 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS, IEEE, 2020, pp. 81–84, https://doi.org/10.1109/M2GARSS47143.2020.9105154.
[13] N. Muthukumar, Few-shot learning text classification in federated environments, in: 2021 Smart Technologies, Communication and Robotics (STCR), IEEE, 2021, pp. 1–3.
[14] X. Li, F. Pu, R. Yang, R. Gui, X. Xu, AMN: attention metric network for one-shot remote sensing image scene classification, Remote Sens 12 (2020) 4046, https://doi.org/10.3390/rs12244046.
[15] D. Argüeso, A. Picon, U. Irusta, A. Medela, M.G. San-Emeterio, A. Bereciartua, A. Alvarez-Gila, Few-Shot Learning approach for plant disease classification using images taken in the field, Comput. Electron. Agric. 175 (2020) 105542.
[16] Y. Li, X. Chao, Semi-supervised few-shot learning approach for plant diseases recognition, Plant. Methods 17 (2021) 1–10, https://doi.org/10.1186/s13007-021-00770-1.
[17] S. Janarthan, S. Thuseethan, S. Rajasegarar, Q. Lyu, Y. Zheng, J. Yearwood, Deep metric learning based citrus disease classification with sparse data, IEEE Access 8 (2020) 162588–162600, https://doi.org/10.1109/ACCESS.2020.3021487.
[18] H.F. Ng, J.J. Lo, C.Y. Lin, H.K. Tan, J.H. Chuah, K.H. Leung, Fruit ripeness classification with few-shot learning, in: Proceedings of the 11th International Conference on Robotics, Vision, Signal Processing and Power Applications, Springer, Singapore, 2022, pp. 715–720.
[19] A. Afifi, A. Alhumam, A. Abdelwahab, Convolutional neural network for automatic identification of plant diseases with limited data, Plants 10 (2020) 28, https://doi.org/10.3390/plants10010028.
[20] Y. Li, J. Yang, Few-shot cotton pest recognition and terminal realization, Comput. Electron. Agric. 169 (2020) 105240, https://doi.org/10.1016/j.compag.2020.105240.
[21] S.V. Nuthalapati, A. Tunga, Multi-domain few-shot learning and dataset for agricultural applications, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1399–1408.
[22] Ooi, A.C. (2015). Common insect pests of rice and their natural biological control.
[23] Y. Ma, F. Li, Self-challenging mask for cross-domain few-shot classification, in: 2022 26th International Conference on Pattern Recognition (ICPR), IEEE, 2022, pp. 4456–4463.
[24] Y. Chen, Z. Liu, H. Xu, T. Darrell, X. Wang, Meta-baseline: exploring simple meta-learning for few-shot learning, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, IEEE, Montreal, QC, 2021, pp. 9062–9071, https://doi.org/10.1109/ICCV48922.2021.00893.
[25] N. Yavari, Few-Shot Learning with Deep Neural Networks for Visual Quality Control, Evaluations on a Production Line, 2020.
[26] Y. Pan, T. Yao, Y. Li, Y. Wang, C.W. Ngo, T. Mei, Transferrable prototypical networks for unsupervised domain adaptation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 2239–2247.
[27] Z. Ji, X. Chai, Y. Yu, Y. Pang, Z. Zhang, Improved prototypical networks for few-shot learning, Pattern Recognit. Lett. 140 (2020) 81–87, https://doi.org/10.1016/j.patrec.2020.07.015.
[28] S. Albawi, T.A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network, in: 2017 international conference on engineering and technology (ICET), Ieee, 2017, pp. 1–6.
[29] Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., & Huang, J.B. (2019). A closer look at few-shot classification. arXiv preprint arXiv:1904.04232.

Further reading

[30] Ball, J. (2021). Few-Shot Learning for Image Classification of Common Flora. arXiv preprint arXiv:2105.03056.
[31] Ilievski, V., Musat, C., Hossmann, A., & Baeriswyl, M. (2018). Goal-oriented chatbot dialog management bootstrapping with transfer learning. arXiv preprint arXiv:1802.00500.
[32] S. Jadon. SSM-Net for Plants Disease Identification in Low Data Regime, 2020 IEEE/ITU International Conference on Artificial Intelligence for Good (AI4G), IEEE, 2020, pp. 158–163.
[33] J. Jiménez-Luna, F. Grisoni, G. Schneider, Drug discovery with explainable artificial intelligence, Nat. Machine Intelli. 2 (2020) 573–584, https://doi.org/10.1038/s42256-020-00236-4.
[34] Y. Li, J. Yang, Meta-learning baselines and database for few-shot classification in agriculture, Comput. Electron. Agric. 182 (2021) 106055, https://doi.org/10.1016/j.compag.2021.106055.
[35] Y. Li, L. Zhang, W. Wei, Y. Zhang. Deep self-supervised learning for few-shot hyperspectral image classification, IGARSS 2020-2020, IEEE, 2020, pp. 501–504.
[36] A. Parnami, M. Lee, Learning from few examples: a summary of approaches to few-shot learning, arXiv preprint arXiv:2203.04291. (2022).
[37] L.M. Tassis, R.A. Krohling, Few-shot learning for biotic stress classification of coffee leaves, Artif. Intell. 6 (2022) 57–67, https://doi.org/10.1016/j.aiia.2022.04.001.
[38] C. Wang, J. Zhou, C. Zhao, J. Li, G. Teng, H. Wu, Few-shot vegetable disease recognition model based on image text collaborative representation learning, Comp. Electronics Agri. 184 (2021) 106098, https://doi.org/10.1016/j.compag.2021.106098.
[39] J. Wang, Y. Zhai. Prototypical siamese networks for few-shot learning, IEEE, 2020, pp. 178–181.
[40] Q. Zhong, L. Chen, Y. Qian. Few-shot learning for remote sensing image retrieval with maml, IEEE, 2020, pp. 2446–2450.