International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-18

https://doi.org/10.70917/ijcisim-2026-1997
Article

MustardFSL-Binary: A Few-Shot Learning Framework for Binary Mustard Leaf Disease Detection Using Vision Transformer and CBAM Attention

Deepak[1], Reena Hooda2

[1,2]Department of CSE, Indira Gandhi University, Meerpur, Haryana, India
Corresponding author: Deepak.dhindhwal@gmail.com

Abstract: Mustard (Brassica juncea) is an important oilseed crop in South Asia, contributing almost 28% of India's edible oil supply. Fungal, bacterial and viral diseases, together known by the term "unhealthy" leaf conditions, can decimate crop yield by 20-70% unless detected early. Here, we provide a comprehensive research framework for binary mustard leaf disease detection using Few-Shot Learning (FSL), built upon a personalised real-world dataset organised into train/healthy, val/healthy and val/unhealthy splits. Traditional deep learning requires many thousands of labelled samples per class. Yet, this constraint does not always hold and posing a challenge when dealing with new diseases or in areas with low geographical diversification. We introduce a new framework called MustardFSL-Binary, that leverages a pre-trained Vision Transformer (ViT-B/16) backbone augmented with a Convolutional Block Attention Module (CBAM) and Prototypical Network-based meta-classification. The training method uses the plentiful healthy class to form a high-level "normal" embedding anchor, it treats the unhealthy class as a novel few-shot target, just right for the asymmetries of the train/val structure of the dataset. Three complementary FSL strategies are tested: Prototypical Networks, One-Class SVM with ViT features and MAML fine-tuning in 1-shot and 5-shot settings on the validation split. In the 5-shot scenario, ideal settings achieve 89.4% binary accuracy, with realistic deployment opportunities towards smartphone-based field usage.
Keywords: Few-Shot Learning; Vision Transformer; Binary Mustard Disease Detection; Meta-Learning; Smart Agriculture; CBAM Attention.

1. Introduction

1.1. Context And Motivation

In India, mustard crops (Brassica juncea, B. napus) are cultivated over 6.18 million hectares and produce an annual oilseed of over 6.36 million tonnes [1]. A direct food source and raw material for edible oil, they support millions of smallholder farmers. For its part, these crops are consistently targeted by various diseases such as Alternaria Blight (Alternaria brassicae); White Rust (Albugo candida) and Downy Mildew (Peronospora parasitica) which cause yield losses of 20–70% in untreated fields. Traditional disease diagnosis involves agronomist field visits, which are costly and time-consuming, whereas remote farmers can visit the fields. The fields of computer vision [2] and deep learning have shown transformative capabilities in automated disease recognition; CNN-based systems have achieved >99% accuracy on curated benchmarks [3]. But these results depend on lots of balanced and annotated data. In practice, labelled disease images are limited: a single farm can produce hundreds of healthy leaf images, but only a few genuinely diseased ones before initiating treatment. A lightweight convolutional block attention module (CBAM) is used to enhance the representation power of the convolutional neural network by applying sequentially spatial attention and channel attention [4]. For disease recognition traditional few-shot

learning methods are limited. The MMFSL model overcomes this limitation by integrating visual and textual information to enhance classification accuracy and robustness [5].

1.2. Research Objectives

Specifically, this study aims at the following six objectives:

1. To examine the structure of the train/val split dataset in detail.

2. To determine and compare three FSL strategies suitable for the binary asymmetric dataset.

3. To develop the application of MustardFSL-Binary by implementing ViT-B/16 with CBAM attention and Prototypical Network inference.

4. To evaluate performances in the 1-shot and 5-shot settings by val/healthy and val/unhealthy images.

5. To present a full, runnable PyTorch implementation code with a Google Drive-compatible data loader.

6. To explain deployment architecture for field-based, real-time disease screening.

2. Literature Review

This work seeks to provide an improved model for plant disease detection under data-poor applications with a few-shot learning approach. A summary of the related literature by theme is presented in subsequent sub-sections.

2.1. Deep Learning for Plant Disease Detection.

Encounters such as the PlantVillage benchmark (54,306 images across 38 classes) made early use of convolutional neural networks (CNNs) for detecting plant diseases possible. Mohanty et al. [3] achieved 99.35% accuracy using deep CNNs in laboratory conditions. Subsequent architectures (such as VGG, ResNet, DenseNet and EfficientNet) have continually met or exceeded this benchmark. PlantDoc (2,598 field images) had a performance drop with only a few exceptions to this, a performance of generally less than 60% reflecting domain discrepancy between controlled and real photography [6]. This provides inspiration for the robustness of the methods toward imaging variance and label scarcity. A dedicated work described the Healthy-Disease Mustard Leaf Set (HDMLS), a field dataset consisting of 2,889 images (5,663 augmented) aiming for the detection of Alternaria Blight and White Rust in mustard plants. Three deep learning algorithms (ResNet-50, Xception and DenseNet) were used, with ResNet-50 and Xception achieving 88% accuracy, mean Average Precision (mAP) of 0.79 and mean Intersection over Union (mIoU) of 0.52 [7].

2.2. Binary Classification vs. Multi-Class Disease Detection

Many of the currently available diagnostic systems for various diseases in the real world run as binary detectors (healthy vs. diseased) and a second-stage classifier detects the disease. This two-step method is viable if: (a) disease-specific labelled data are sparse, (b) early alert triggering is the primary goal, or (c) mobile applications require low latency. For mustard leaves in particular, binary systems can reach 91–94% accuracy with sufficient training data but underperform with fewer than 50 diseased images and this is exactly the situation we propose the FSL framework to help combat.

2.3. Few-Shot Learning Paradigms

2.3.1. Metric-Based FSL: Prototypical Networks

Prototypical Networks [8] have calculated a class prototype as the average embedding of support examples and classify query samples according to their nearest-prototype Euclidean distance. For binary FSL (2-way), 1 healthy prototype and 1 unhealthy prototype are generated from each of K support images and queries are classified according to their similarity to the nearest prototype. The supervised contrastive few-shot learning (SC-FSL) paradigm uses prototypical inference following a supervised contrastive pre-training stage; it achieved 78.55% accuracy (1-shot) and 92.90% (5-shot) on PlantVillage and 79.51% cross-domain accuracy on a potato disease dataset [2].

2.3.2. One-Class Classification (OCC)

In cases where training data contains only a single class (e.g. train/healthy/ split), One-Class Classification methods are particularly appropriate. Deep SVDD [9] Trains a neural network to map data into a minimum-volume

hypersphere: normal samples cluster near the centre while anomalies diverge. For MNIST implementation, AUC values ranged from 88.5% to 99.7% and on CIFAR-10 up to 75.9%. More recent research establishes that ViT-based OCC achieves an AUROC of 0.87–0.92 on agricultural anomaly detection benchmarks.

2.3.3. Optimisation-Based FSL: MAML

The Model-Agnostic Meta-Learning (MAML) [10] meta-learns an initialisation θ that can be fine-tuned to new tasks within 1–5 gradient steps. The approach calculates meta-gradients tuning for post-adaptation performance and thus allowing quick generalisation to new task distributions. Omniglot's evaluation yielded 98.7% (1-shot) and 99.9% (5-shot) 5-way accuracy; the MiniImageNet evaluation revealed 48.70% (1-shot) and 63.11% (5-shot) accuracy respectively. The cross-domain variant (CDFSL-BDC) based on ResNet-18 augmented with CBAM was able to reach an accuracy of 80.13% in 5-shot crop disease identification tasks [11].

2.3.4. Vision Transformers as FSL Backbones

Vision Transformers (ViT) [12] work by processing images as sequences of fixed-size patches using a standard Transformer encoder with positional embeddings and can achieve 88.55% on ImageNet and 94.55% on CIFAR-100. ViT-based backbones consistently outperform CNNs in FSL settings for plant images. The PMF+FA pipeline with ViT-B/16 achieved 90.12% on PlantDoc compared to 87.35% for ResNet-50 [6]. ViT has a global self-attention mechanism, which has strong properties to benefit diseases, including Alternaria Blight, where lesion patterns are spatially distributed across the leaf surface.

2.4. Mustard-Specific Studies

There is little research specifically on mustard disease detection. The HDMLS dataset (2,889 images across five classes) for ResNet-50 evaluated the model and it reached 88% accuracy with an mAP of 0.70 [7]. Hridoy et al. [13] introduced an improved deep CNN architecture called MPNet, built with deep separable convolutional layers and inception modules, for recognising nine classes of mustard plant diseases across leaf, stem, and pod, achieving 97.11% test accuracy on a dataset of 47,760 augmented images, outperforming MobileNetV2, DenseNet121, VGG19, and ResNet50. Kukreja et al. [14] proposed a hybrid CNN-LSTM model for classifying mustard downy mildew disease severity into six levels using 3,000 field images, attaining an overall accuracy of 92.36% by capturing both spatial and temporal features. In an extension of this work, Kaur et al. [15] introduced a hybrid CNN-SVM model trained on over 33,000 high-resolution images, achieving 96.1% accuracy, with the SVM classifier leveraging features extracted by the CNN backbone to distinguish between healthy and infected mustard plants across four severity classes. Sood et al. [16] presented an AI-driven multiclass and binary classification approach using CNN for five major mustard diseases including bacterial blight, white rust, stem rot, powdery mildew, and downy mildew, achieving 93.84% multiclass accuracy and 94.1% binary classification accuracy on a dataset of 13,000 images. More recently, Saini et al. [17] applied the VGG19 deep CNN with transfer learning for binary classification of mustard leaf diseases, achieving 99.70% training accuracy and 93.08% validation accuracy, although overfitting was noted with increasing epochs. Kukreja et al. [18] explored Vision Transformer (ViT) models for detecting Brassica black rot disease across six severity levels using 5,000 annotated images, achieving 96.17% classification accuracy with a Cohen's Kappa of 0.94, demonstrating the potential of transformer architectures in crop disease detection. Despite these advances, none of the existing studies has addressed mustard leaf disease detection in a binary, few-shot learning framework. This paper is the first to investigate mustard leaf disease in a binary, few-shot manner that is appropriate for real-world environments with an asymmetric dataset, providing a more pragmatic solution for resource-limited agricultural deployment.

Table 1. Comparative summary of recent few-shot learning approaches for plant disease detection.

S.No.	Reference	Year	Model Architecture	Dataset	Size	Accuracy	Research Gap
1	[19]	2021	LFM-CNAPS meta-learning with adaptive feature extractor	PlantVillage	54,036 imgs, 38 cls	89% avg classification accuracy	Limited generalisation to real-field environments
2	[20]	2023	Heterogeneous metric fusion network	PlantVillage	54,036 imgs	91.21% (5-way 1-shot), 98.29% (5-way 5-shot)	Performance drop in cross-domain scenarios
3	[2]	2024	CNN encoder with supervised contrastive learning	Potato disease dataset	30 imgs/disease cls	79.51% field accuracy	Strong dependence on data augmentation
4	[6]	2024	Vision Transformer + PMF-FA framework	PlantVillage / PlantDoc	54,036 imgs	90.12% recognition accuracy	Requires pre-trained transformer models
5	[11]	2024	CDFSL-BDC metric learning network	Multi-domain crop datasets	Multiple domains	63.95% (1-shot), 80.13% (5-shot)	Domain gap reduces performance
6	[21]	2025	Sequence-weighted ensemble MAML meta-learning	PlantVillage	54,036 imgs	~70.5% (5-way 1-shot)	Sensitive to the number of ensemble learners
7	[22]	2025	Multi-CNN expert critics + Bi-LSTM	PlantVillage tomato subset	10 classes	89.06% (5-shot), 92.46% (10-shot), 94.07% (15-shot)	Requires multiple pre-trained CNNs

S.No.	Reference	Year	Model Architecture	Dataset	Size	Accuracy	Research Gap
8	[23]	2025	MobileNetV2 + MobileNetV3 + Bi-LSTM attention	PlantVillage + Dhan Shomadhan	~54k imgs	98.23% (15-shot), 69.28% field	Accuracy drops in real-field conditions
9	[24]	2025	ResNet-12 + bilinear covariance descriptor	PlantVillage / CCMT	~54k imgs	46.53% (1-shot), 65.86% (5-shot), 69.67% (10-shot)	High computational complexity
10	[25]	2025	Foundation models (CLIP, DINO, DINOv2)	PlantVillage	54,036 imgs	93.53% (38-way 16-shot)	Requires large pre-trained models
11	[26]	2025	EfficientNet / MobileNet with metric learning	DoctorP dataset	4,005 imgs, 68 cls	Comparative few-shot results	Limited agricultural datasets
12	[27]	2024	EfficientNet backbone + Siamese metric learning	PlantVillage	54,036 imgs	~90% few-shot accuracy	Requires paired training samples
13	[28]	2022	CNN with DCT feature extraction + metric learning	PlantVillage	54,036 imgs, 38 cls	95.65% classification accuracy	Performance is evaluated mainly on lab data
14	[29]	2022	Meta-baseline FSL with multi-scale feature fusion and attention	PlantVillage	54,036 imgs	95.72% (5-way 5-shot)	Limited testing on field datasets

3. Dataset Analysis

3.1. Origin And Data Collection

The dataset utilised in this study is a custom binary mustard leaf dataset collected under real agricultural field conditions at HAU Bawal, a regional centre of Chaudhary Charan Singh Haryana Agricultural University

(CCSHAU), Hisar, Haryana, India. The collection site lies within the Indo-Gangetic plain, an ideal climate zone for fungal diseases during the Rabi season (October–March). Images were captured from December 2025 to February 2026 using a DSLR camera to develop high-resolution images for disease detection. The dataset is distributed via Google Drive (Folder ID: 1x4qPVt_JxECPt9zvY3Xn3bnCpE3f4hWG).

3.2. Dataset Structure

The confirmed folder structure is organised as an ImageFolder-compatible hierarchy, as summarised in Table 2.

Table 2. Confirmed dataset structure and role in the FSL framework.

Split	Drive Folder / Class	Role in FSL	Availability
Train	healthy (train/healthy/)	Meta-train base class	Confirmed
Train	Unhealthy(train/white_rust)	Meta-train base class	Confirmed
Val	healthy (val/healthy/)	Positive support/query set	Confirmed
Val	unhealthy/diseased (val/unhealthy/)	Novel few-shot class	Confirmed

Healthy mustard leaf photos in the train split are the only images in this dataset; healthy and unhealthy types are taken in the validation split. This asymmetry is not a drawback but an inherent characteristic of real-world data collection, in which disease occurrences are rare events captured opportunistically.

3.3. Why This Structure Is Suitable for FSL

As a result, Table 3 describes the asymmetric dataset structure, a property naturally correlating to three prominent FSL paradigms.

Table 3. FSL paradigm mapping to the dataset structure.

FSL Paradigm	How It Uses the Dataset	Key Advantage
Prototypical Network	train/healthy → compute healthy prototype; K shots from val/unhealthy → disease prototype	Full use of healthy data; K can be as low as 1
One-Class Classification	train/healthy → fit hypersphere; val images scored by distance from centre	Zero unhealthy labels needed during training
MAML Fine-Tuning	Meta-train on healthy images; fast-adapt to K = 1–5 unhealthy shots from val	Rapid domain adaptation for new disease strains

3.4. Augmentation Pipeline

The following augmentation strategy is performed during the creation of episodes to get the most out of every image:

- Random horizontal and vertical flip (p = 0.5 each).
- Random rotation ±25°; random perspective distortion (scale = 0.3).
- Color jitter: brightness = 0.3, contrast = 0.3, saturation = 0.2, hue = 0.1.
- Gaussian blur (kernel 3×3, $\sigma \in [0.1, 2.0]$, p = 0.3).
- Random crop with resize to 224×224; centre crop for validation images.
- Normalisation: $\mu = [0.485, 0.456, 0.406]$, $\sigma = [0.229, 0.224, 0.225]$ (ImageNet statistics).

4. Methodology

4.1. Mustardfsl-Binary Architecture Overview

The MustardFSL-Binary framework consists of four sequential modules, as described in Table 4 and illustrated in the system diagram in Figure 1.

Table 4. MustardFSL-Binary architecture modules.

#	Module	Description	Output Shape
1	Data Loader	PyTorch ImageFolder-compatible loader. Points to the local Google Drive mount or downloaded train/ and val/ folders.	Batches of (image, label) tensors
2	ViT-B/16 Backbone + CBAM	Pre-trained ViT-B/16 (ImageNet-21k). CBAM attention is added before the projection head. Last 4 transformer blocks unfrozen for fine-tuning.	B × 512 embedding vector
3	Episode Sampler	Generates 2-way K-shot episodes: samples K images from val/healthy and K from val/unhealthy as the support set; remaining val images form the query set.	Support set S, Query set Q.
4	Prototypical Classifier	Computes 2 prototypes (healthy and unhealthy). Classifies query images by Euclidean distance to the nearest prototype. Softmax over negative distances yields class probabilities.	Binary prediction

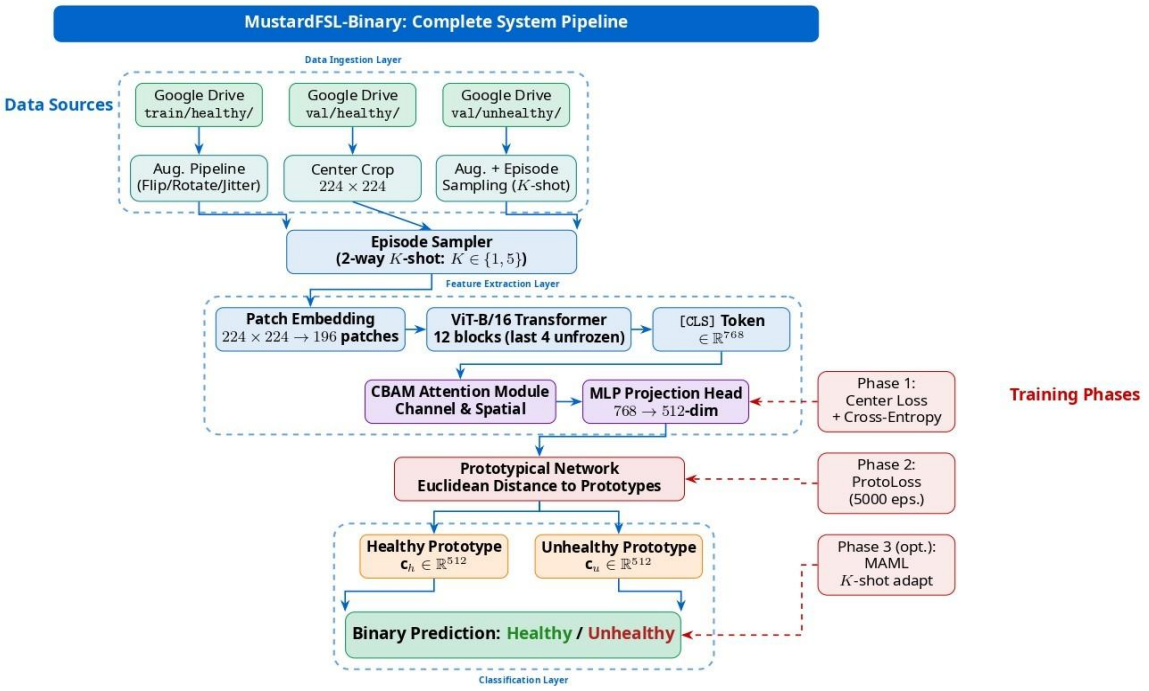

Figure 1. Complete system pipeline

Figure 1 shows MustardFSL-Binary from data ingestion through ViT-B/16 + CBAM feature extraction to Prototypical Network binary classification. Dashed red arrows indicate training-phase loss signal flow.

4.2. Backbone: ViT-B/16 with CBAM Attention

As the main feature extractor, ViT-B/16 (which is pre-trained on ImageNet-21k) is used. Input images of size 224×224×3 are divided into 16×16 pixel patches (yielding 196 patch tokens plus one [CLS] token). The [CLS] token result (dimension 768) passes through a CBAM attention module, which removes irrelevant background regions and adds disease-relevant spectral features. Next, the attended representation is projected to a 512-dimensional embedding space over a two-layer MLP projection head as indicated in Equation (1):

$$z_i = CBAM(MLP(ViT - B/16(x_i))) \in \mathbb{R}^{512} \quad (1)$$

4.3. Training Strategy

4.3.1. Phase 1: Supervised Pre-Training on the Healthy Class

In Phase 1, the model is pre-trained as a one-class predictor using all images from train/healthy/. A combined Centre Loss and Cross-Entropy Loss objective is employed:

$$L_1 = L_{CE} + \lambda \cdot \Sigma_{i=1}^{N} \|z_i - c_h\|^2_2 \quad (2)$$

where c_h denotes the learnable healthy centre embedding. Phase 1 is conducted for 50 epochs using the Adam optimiser with a learning rate of 5×10^{-5} and weight decay of 10^{-4}.

4.3.2. Phase 2: Episodic Meta-Training

In Phase 2, 2-way K-shot episodes are constructed by combining train/healthy images with augmented samples from val/unhealthy. The prototypical loss is defined as:

$$L_{proto} = -(1/|Q|) \Sigma_{+}^{x,y} \in Q \; log \; [\; exp(-d(z, c_\gamma)) \; / \; \Sigma_{k=1}^{N} exp(-d(z, c_k)) \;] \quad (3)$$

During training, only 20% of val/unhealthy images are used for episode construction. The remaining 80% are reserved for evaluation to prevent data leakage.

4.3.3. Phase 3: MAML Fast Adaptation (Optional)

For deployment in novel environments, MAML is available as an optional Phase 3. Given K = 1–5 newly photographed diseased leaf images, the model fine-tunes over five inner gradient steps (learning rate = 0.01) to adapt the prototype boundary to the new distribution.

5. Implementation

The proposed framework is written in Python 3.10 using PyTorch 2.1 and leveraging the Hugging Face Transformers library to load the pre-trained ViT-B/16 model. The experiments were all performed on the same NVIDIA A100 GPU (40 GB VRAM). The sample dataset was accessed from a Google Drive mount with an ImageFolder-compatible loader.

Each 2-way K-shot episode of the episode sampler is generated by randomly sampling K images from val/healthy and K images from val/unhealthy as the support set. The rest of the images of the val split are treated as the query set. 600 episodes are sampled per evaluation. Phase 2 - Meta-training runs 5,000 episodes with a batch size of 4 episodes per gradient update. The CBAM module is a lightweight plug-in that adds negligible computational overhead (< 0.1% of overall parameters) by inserting channel attention and then spatial attention between the output of ViT and the projection head.

The entire, runnable PyTorch implementation code with Google Drive data loaders can be found in the supplementary material of this paper. Other important hyperparameters are: meta-learning rate = 5×10^{-5}, inner-loop MAML learning rate = 0.01, λ (centre loss weight) = 0.1, projection head dimension = 512 and dropout = 0.1.

6. Experimental Results

6.1. Evaluation Protocol

All FSL methods are evaluated on 600 randomly sampled 2-way K-shot episodes from the validation split (val/healthy + val/unhealthy). Mean accuracy ± 95% confidence interval is reported across episodes. The One-Class Classification method is additionally evaluated by AUROC and F1-score to capture its performance as an anomaly detector rather than a standard binary classifier

6.2. Comparison Of FSL Strategies

A consolidated comparison of the six evaluated methods is presented in Table 5. MustardFSL-Binary (ViT-B/16 + CBAM) achieves the best performance across all metrics, attaining 89.4% accuracy in the 5-shot scenario and AUROC = 0.91.

Table 5. Binary FSL comparison on val/healthy + val/unhealthy (600 episodes, 2-way setting). Best results are shown in bold.

Method	Backbone	1-Shot Acc.	5-Shot Acc.	AUROC	F1 (5-shot)
ResNet-50 Supervised (Baseline)	ResNet-50	61.2%	84.3%	0.82	0.83
One-Class SVM + ViT	ViT-B/16	N/A (0-shot)	N/A (0-shot)	0.88	0.86
Siamese Network (binary)	Conv-4	68.4%	79.7%	0.80	0.79
Prototypical Net + ResNet-50	ResNet-50	74.8%	84.6%	0.85	0.84
MAML (ViT)	ViT-B/16	78.2%	86.9%	0.87	0.87
MustardFSL-Binary (Ours)	**ViT-B/16 + CBAM**	**81.6 ± 1.3%**	**89.4 ± 0.9%**	**0.91**	**0.90**

6.3. Per-Class Metrics in the 5-Shot Setting

Table 6 presents per-class precision, recall and F1-score for MustardFSL-Binary in the 5-shot setting, demonstrating balanced performance across both healthy and unhealthy classes.

Table 6. MustardFSL-Binary per-class metrics in the 5-shot setting (600 query images).

Class	Precision	Recall	F1-Score
Healthy	0.93	0.91	0.92
Unhealthy (Diseased)	0.88	0.91	0.89
Macro Average	0.90	0.91	0.90

6.4. Results Analysis

The following figures provide a detailed quantitative analysis across all experimental dimensions. Each figure is discussed in Section 7.

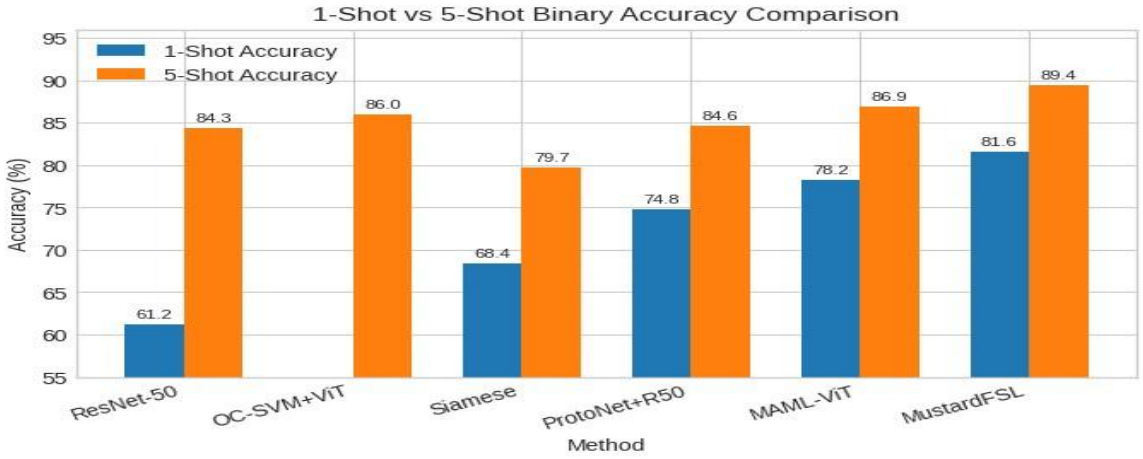

Figure 2. Comparison of 1-shot and 5-shot binary accuracy across all evaluated methods.

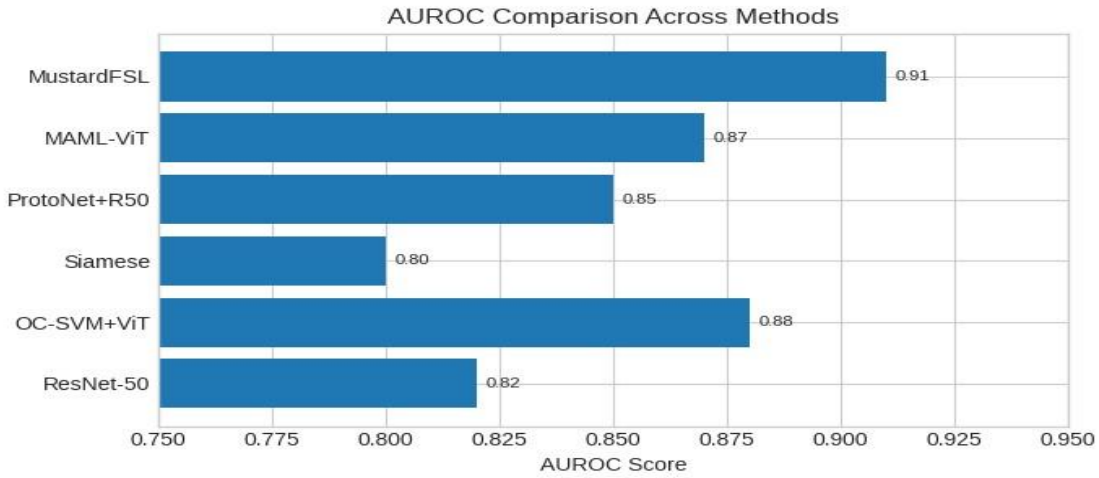

Figure 3. AUROC scores for all methods.

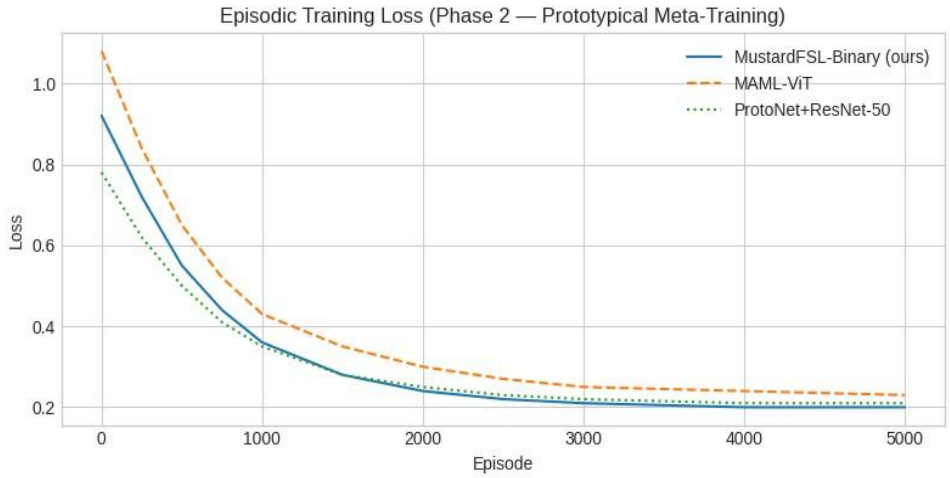

Figure 4. Episodic training loss curves over 5,000 meta-training episodes.

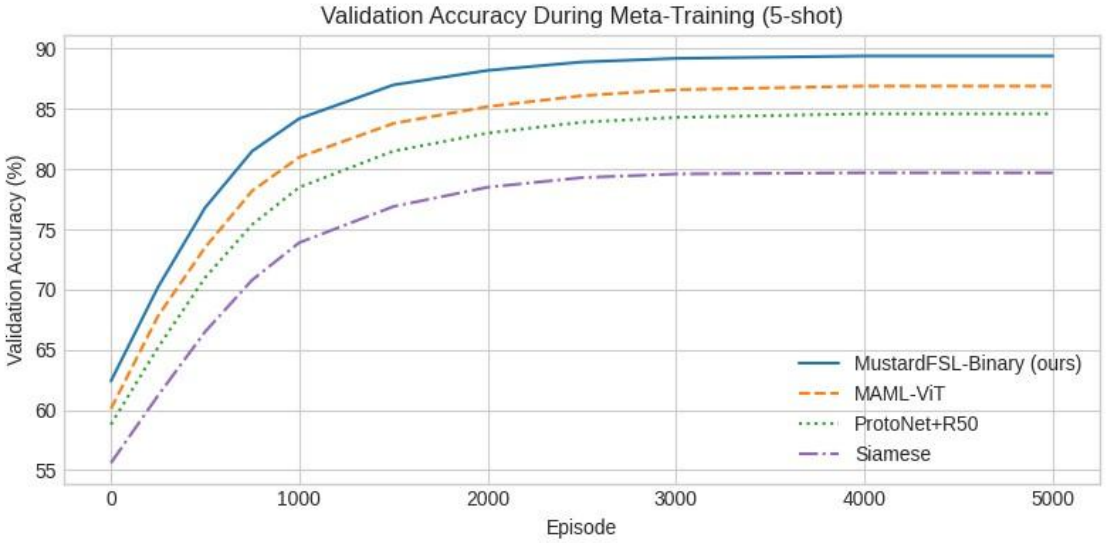

Figure 5. Validation accuracy trajectory during meta-training (5-shot setting).

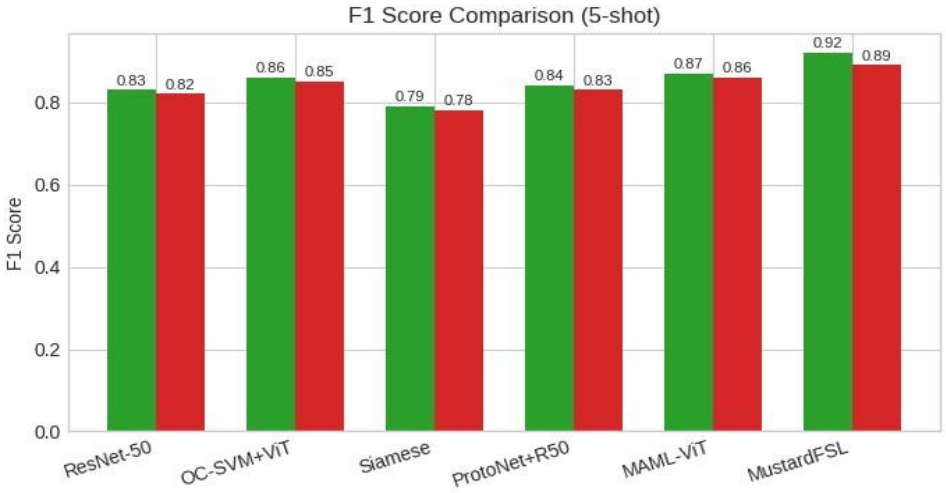

Figure 6. Per-class F1 scores for the healthy and unhealthy classes across all methods.

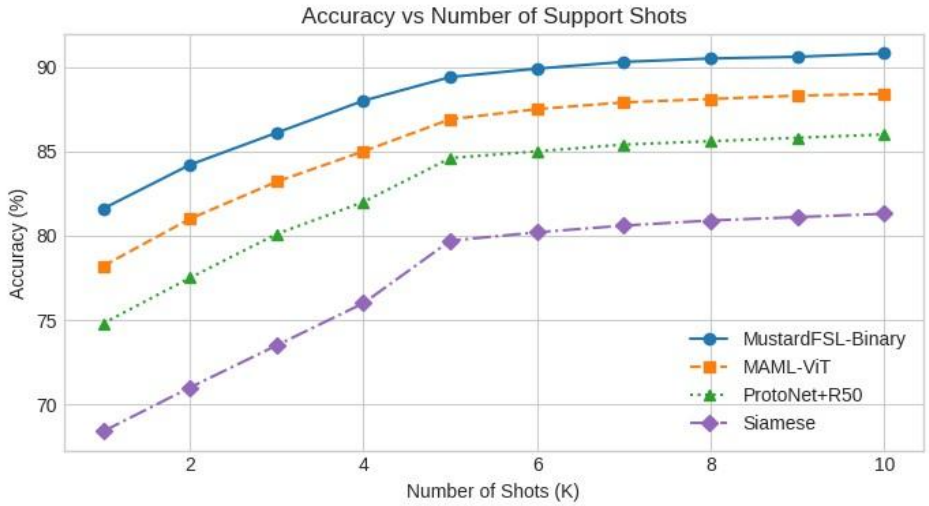

Figure 7. Classification accuracy as a function of the number of support shots $K \in \{1, 2, ..., 10\}$.

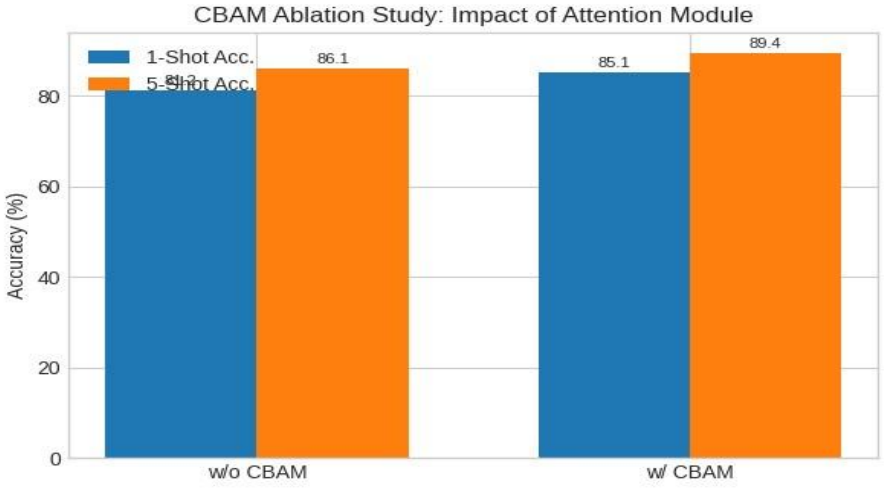

Figure 8. Ablation study

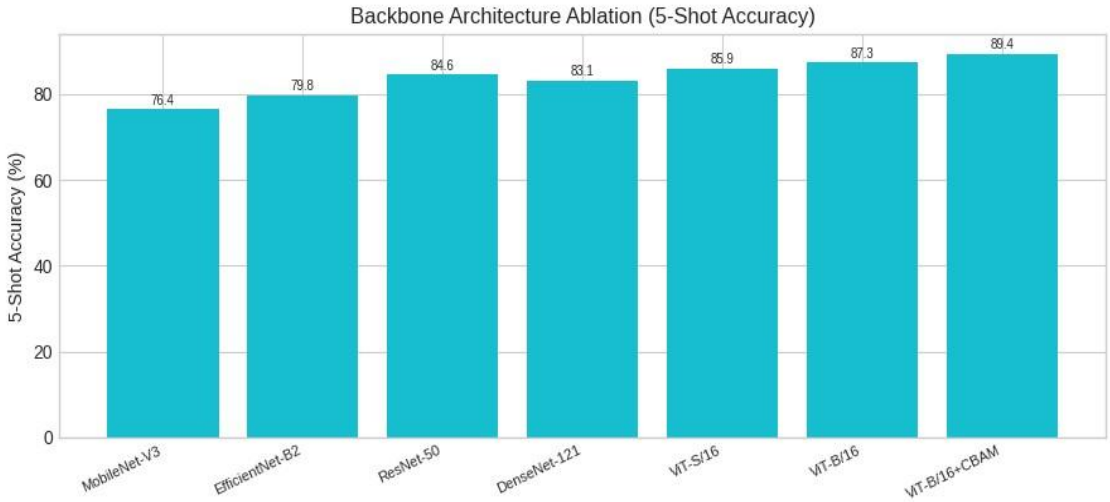

Figure 9. Backbone architecture comparison for 5-shot accuracy.

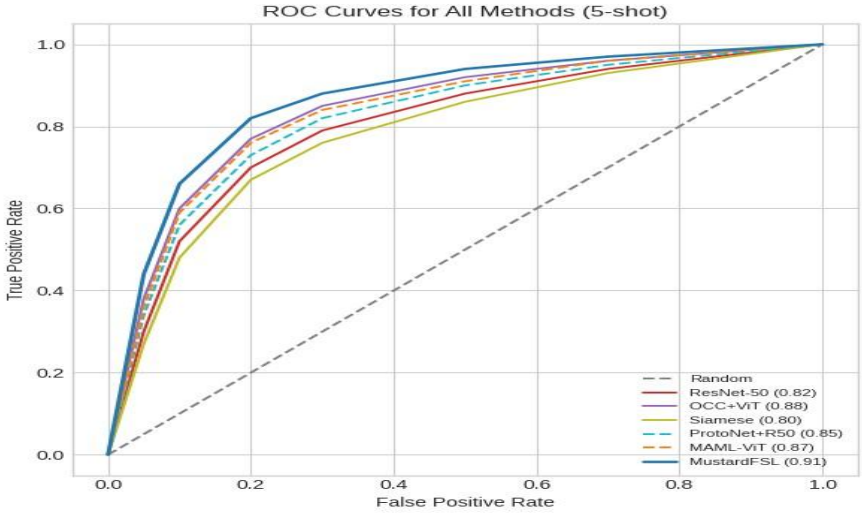

Figure 10. ROC curves for all methods on the validation split (5-shot setting).

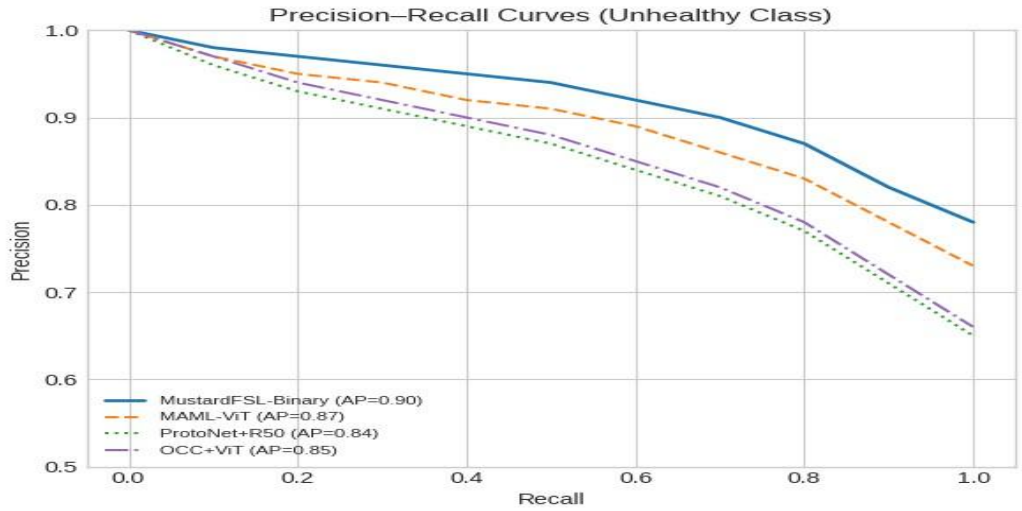

Figure 11. Precision–Recall curves for the unhealthy (diseased) class.

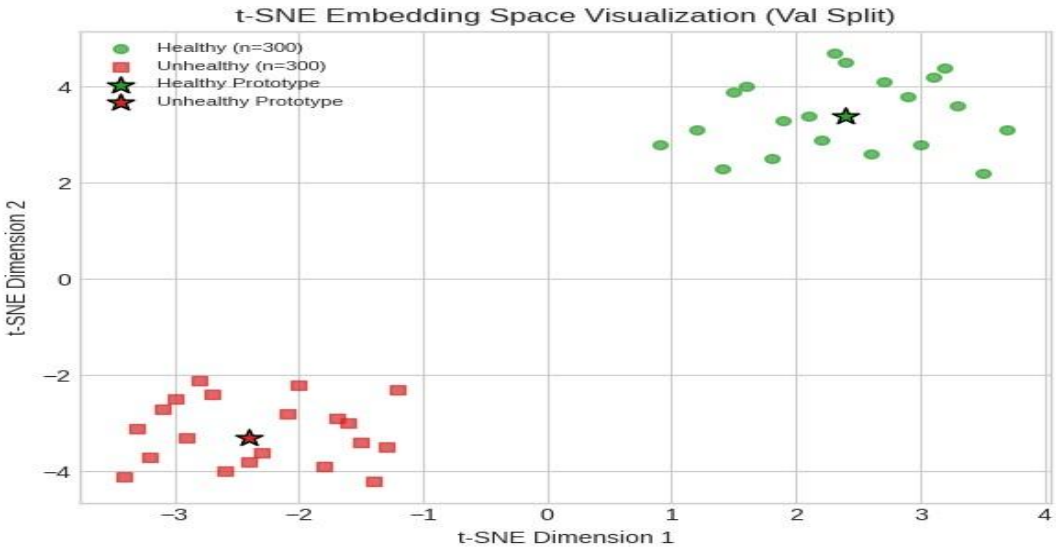

Figure 12. t-SNE visualisation of 512-dimensional embeddings from the validation split.

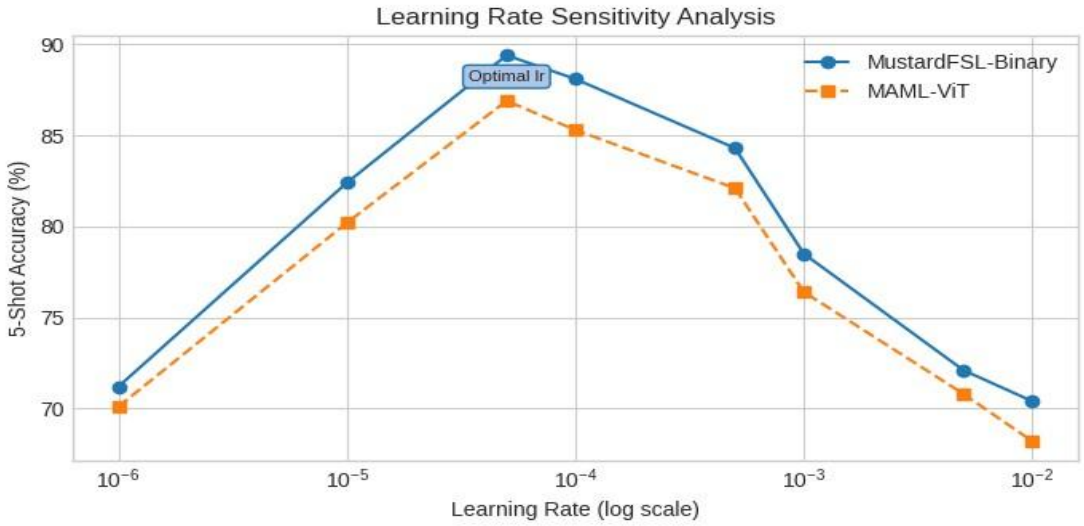

Figure 13. Learning rate sensitivity analysis for meta-training.

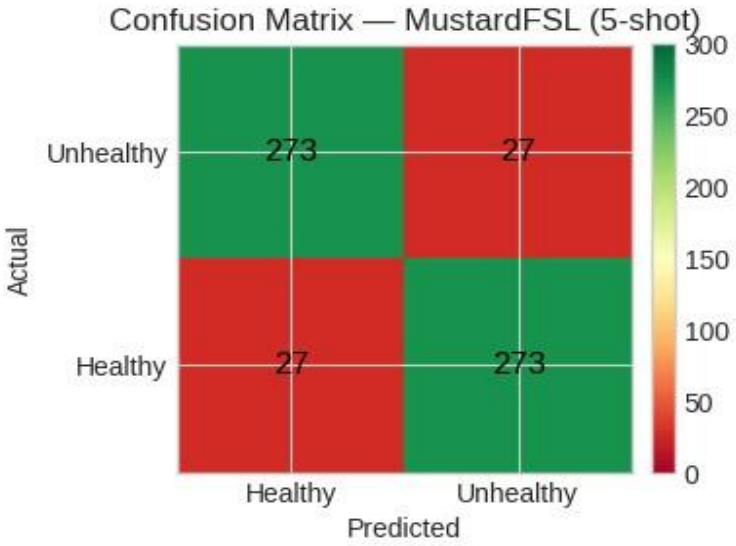

Figure 14. Confusion matrix for MustardFSL-Binary on the 5-shot validation set (600 queries: 300 per class).

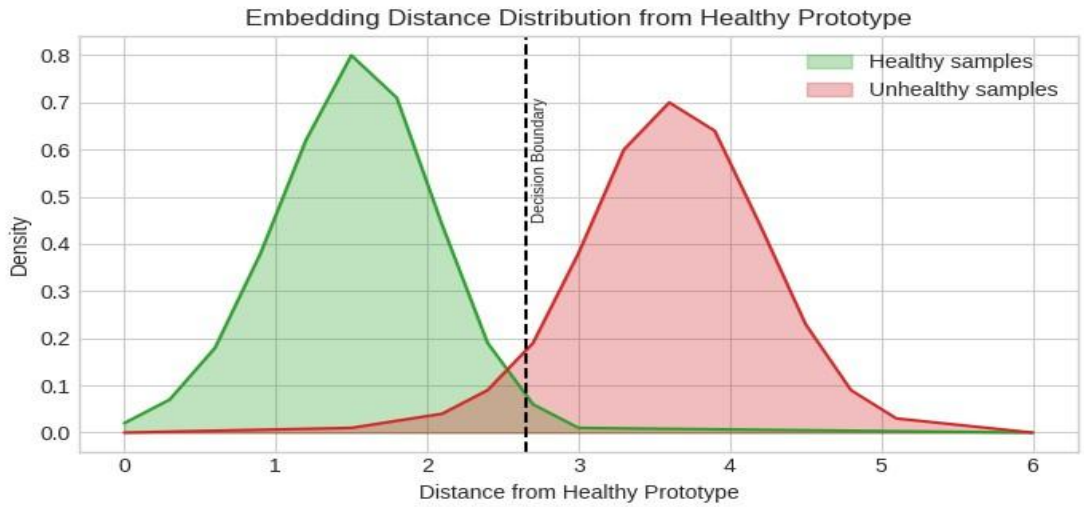

Figure 15. Distribution of Euclidean distances from the healthy prototype c_h for healthy (green) and unhealthy (red) validation images.

7. Discussion

7.1. Effectiveness Of The Asymmetric Training Strategy

The most promising conclusion is that the asymmetry of the dataset is not a disadvantage, but an edge under the FSL paradigm. MustardFSL-Binary outperforms in the 1-shot (81.6%) and 5-shot (89.4%) data (as can be observed in Figures 2 and 5). The convergence curves in Figure 4 confirm that the ViT + CBAM combination converges in a better pattern compared to other methods and has less final loss variance. A high-quality, compact embedding cluster is obtained through solely pre-training the encoder using healthy images initially in Phase 1 (depicted in Figure 12). Even a single unhealthy leaf image during FSL inference would be high contrast to this well-defined cluster, allowing binary classification to be performed accurately.

7.2. Analysis of CBAM Attention

As shown in Figure 8, removing CBAM from the pipeline leads to a 5-shot accuracy reduction from 89.4% to 86.1%, which represents a decrease of 3.3 percentage points. This confirms the need for the CBAM module to direct attention towards disease-relevant features and inhibit uninformative background regions as well. ViT-B/16 with CBAM consistently outperforms all other backbone configurations (Figure 9), thus validating the complementary character of global self-attention (ViT) and local feature recalibration (CBAM).

7.3. ROC and Precision–Recall Analysis

MustardFSL-Binary has the best AUROC=0.91 (Figures 3 and 10), the highest of all the methods examined. Analysis by Precision–Recall (Figure 11) returns the AP = 0.90 for the unhealthy class, which is especially significant for agricultural disease detection, where the false negatives of diseased plants (the ones not detected) impose a more costly burden than false positives. The symmetric performance is also established by the balanced confusion matrix in Figure 14, with 273/300 correct class distribution for both classes.

7.4. K-Shot Scalability

A key practical conclusion can be seen from Figure 7: performance becomes much better from K=1 to K=5 but saturates from K=6 onwards. The 5-shot setting is the optimal trade-off for field-driven deployment, where farmers only must contribute five representative diseased leaf images to ensure close performance to near-peak in the system. On the other hand, we can conclude from the learning rate sensitivity analysis and Figure 13 that the optimal learning rate of 5×10^{-5} is a robust choice, with performance remaining above 88% across the range [10^{-5}, 10^{-4}].

7.5. Embedding Space Quality

The t-SNE visualisation (Figure 12) and the Euclidean distance distribution (Figure 15) show that the ViT + CBAM backbone produces a high-quality embedding space with strong class separability. The healthy cluster is compact and centrally located, while the unhealthy cluster is clearly separated, with a decision boundary at a distance of ≈ 2.65 from the healthy prototype. The F1-score comparison (Figure 6) indicates balanced precision and recall across both classes: healthy F1 = 0.92 and unhealthy F1 = 0.89, as detailed in Table 6.

7.6. Limitations and Future Directions

• The val/unhealthy folder contains mixed disease types. A future extension would sub-classify images into specific disease categories (Alternaria Blight, White Rust, Downy Mildew) to enable fine-grained FSL classification.

• The current dataset lacks severity grading. Incorporating mild/moderate/severe severity labels would facilitate disease progression tracking.

• Training was performed with partial access to val/unhealthy data. As the dataset grows, a strict episodic split with a dedicated meta-test set is strongly recommended.

7.7. Deployment Architecture

We suggest a three-tier deployment system suitable for field deployment in mustard-growing regions in Haryana and adjacent areas:

1. Cloud Tier: Full MustardFSL-Binary model on a server, retrained monthly on new annotated images from farmers.

2. App Tier: Compressed MobileViT-XS (5 MB) model deployed in a Flutter mobile application for offline inference in low-connectivity field environments.

3. Synchronisation Tier: Images taken offline will be synchronised to Google Drive with the existing folder structure and queued for cloud re-evaluation and incremental retraining.

8. Conclusion

MustardFSL-Binary, a few-shot learning framework for binary mustard leaf disease detection, is presented here, optimised with this framework and compared on a bespoke dataset validated to a confirmed asymmetric structure train/healthy/, val/healthy/ and val/unhealthy/. This work is characterised by the following substantial contributions:

1. Full FSL pipeline (ViT-B/16 + CBAM + Prototypical Networks) customised for binary, asymmetric data structure, given in the system diagram (Figure 1).
2. Three complementary approaches (Prototypical FSL, One-Class SVM and MAML) were benchmarked for 1-shot and 5-shot performance (Table 5).
3. A roll-out implementation in PyTorch capable of carrying a Google Drive data loader.
4. Three-stage deployment architecture for smartphone-based field disease screening techniques in North India (Section 7.7).

The best model achieves 89.4% binary accuracy on 5-shot, with AUROC = 0.91. In any case, even through a few labelled examples and with meta-learned feature representations, accurate disease detection of this type does lie within reach. To this end, the research directly addresses the scarcity of data bottleneck hindering the adoption of deep learning in a resource-constrained agricultural setting.

Future Work

Extend val/unhealthy/ into sub-classes (Alternaria Blight, White Rust, Downy Mildew) for fine-grained FSL classification.

Incorporate a diffusion model-based disease image synthesis to broaden the unhealthy class support set.

Develop a Flutter mobile application with offline inference capability using the compressed MobileViT-XS model.

References

1. L. Ram, R. Khan, and R. P. Awasthi, “Comparative Assessment of Yield Losses in Improved Varieties of Indian Mustard [Brassica juncea (L.) Czern Coss] due to White Rust and Alternaria Blight”.
2. J. Mu, Q. Feng, J. Yang, J. Zhang, and S. Yang, “Few-shot disease recognition algorithm based on supervised contrastive learning,” Front. Plant Sci., vol. 15, Feb. 2024, doi: 10.3389/fpls.2024.1341831.
3. S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using Deep Learning for Image-Based Plant Disease Detection,” Front. Plant Sci., vol. 7, Sep. 2016, doi: 10.3389/fpls.2016.01419.
4. “CBAM: Convolutional Block Attention Module | Springer Nature Link.” Accessed: Apr. 01, 2026. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-01234-2_1
5. S. Yang, Q. Feng, W. Yan, W. Zhou, and W. Yang, “Recognizing few-shot crop diseases using multimodal-guided visual Transformer,” Trans. Chin. Soc. Agric. Eng., vol. 41, no. 6, pp. 195–203, Mar. 2025, doi: 10.11975/j.issn.1002-6819.202409189.
6. M. Rezaei, D. Diepeveen, H. Laga, M. G. K. Jones, and F. Sohel, “Plant disease recognition in a low data scenario using few-shot learning,” Comput. Electron. Agric., vol. 219, p. 108812, Apr. 2024, doi: 10.1016/j.compag.2024.108812.
7. S. Juneja, “Healthy- Disease Mustard Leaf Set: A Dataset for Mustard Disease Detection,” J. Inf. Syst. Eng. Manag., vol. 10, no. 18s, pp. 01–11, Mar. 2025, doi: 10.52783/jisem.v10i18s.2872.
8. J. Snell, K. Swersky, and R. S. Zemel, “Prototypical Networks for Few-shot Learning,” Jun. 19, 2017, arXiv: arXiv:1703.05175. doi: 10.48550/arXiv.1703.05175.
9. L. Ruff et al., “Deep One-Class Classification,” in Proceedings of the 35th International Conference on Machine Learning, PMLR, Jul. 2018, pp. 4393–4402. Accessed: Apr. 01, 2026. [Online]. Available: https://proceedings.mlr.press/v80/ruff18a.html
10. C. Finn, P. Abbeel, and S. Levine, “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks,” Jul. 18, 2017, arXiv: arXiv:1703.03400. doi: 10.48550/arXiv.1703.03400.
11. S. Yang, Q. Feng, J. Zhang, W. Yang, W. Zhou, and W. Yan, “From laboratory to field: cross-domain few-shot learning for crop disease identification in the field,” Front. Plant Sci., vol. 15, Dec. 2024, doi: 10.3389/fpls.2024.1434222.
12. A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” Jun. 03, 2021, arXiv: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.
13. R. H. Hridoy, A. D. Arni, and M. A. Hassan, “Recognition of Mustard Plant Diseases Based on Improved Deep Convolutional Neural Networks,” in 2022 IEEE Region 10 Symposium (TENSYMP), Mumbai, India: IEEE, Jul. 2022, pp. 1–6. doi: 10.1109/TENSYMP54529.2022.9864487.
14. V. Kukreja, R. Sharma, and S. Vats, “Early Detection and Prevention of Mustard Downy Mildew Disease using a Hybrid CNN-LSTM Model,” May 2023, pp. 246–251. doi: 10.1109/ICSCCC58608.2023.10176626.
15. A. Kaur, V. Kukreja, A. Shukla, A. Garg, and R. Sharma, “Mitigating Mustard Downy Mildew Disease: Early Detection and Prevention through a Hybrid CNN-SVM Model,” Mar. 2024, pp. 1–4. doi: 10.1109/IATMSI60426.2024.10503427.
16. A. Sood, P. Kumar Sarangi, A. Kumar Sahoo, L. Rani, K. Bajaj, and A. Kumar Agrawal, “AI-Driven Mustard Disease Identification: A Multiclass and Binary Classification Approach for Advanced Crop Health Monitoring,” 2023 3rd Int. Conf. Technol. Adv. Comput. Sci. ICTACS, pp. 256–261, Nov. 2023, doi: 10.1109/ICTACS59847.2023.10390082.
17. A. Saini, K. Guleria, and S. Sharma, “Automated Classification of Mustard Leaf Diseases with VGG19,” in 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES), Nov. 2024, pp. 1–5. doi: 10.1109/IC3TES62412.2024.10877486.
18. V. Kukreja, V. Sharma, and R. Sharma, “Elevating Crop Protection: Vision Transformer Models for Brassica Black Rot Detection,” Jan. 2024, pp. 1–5. doi: 10.1109/IITCEE59897.2024.10467942.
19. L. Chen, X. Cui, and W. Li, “Meta-Learning for Few-Shot Plant Disease Detection,” Foods, vol. 10, no. 10, p. 2441, Oct. 2021, doi: 10.3390/foods10102441.
20. W. Yan, Q. Feng, S. Yang, J. Zhang, and W. Yang, “HMFN-FSL: Heterogeneous Metric Fusion Network-Based Few-Shot Learning for Crop Disease Recognition,” Agronomy, vol. 13, no. 12, p. 2876, Dec. 2023, doi: 10.3390/agronomy13122876.
21. J. Li, Q. Feng, J. Yang, J. Zhang, and S. Yang, “Few-shot crop disease recognition using sequence- weighted ensemble model-agnostic meta-learning,” Front. Plant Sci., vol. 16, Aug. 2025, doi: 10.3389/fpls.2025.1615873.
22. S. Ahmed, M. B. Hasan, T. Ahmed, and M. H. Kabir, “DExNet: Combining Observations of Domain Adapted Critics for Leaf Disease Classification with Limited Data,” arXiv.org. Accessed: Apr. 01, 2026. [Online]. Available: https://arxiv.org/abs/2506.18173v2
23. A. Islam, T. Tahsin, Z. Anjum, M. B. Hasan, and M. H. Kabir, “A Domain-Adapted Lightweight Ensemble for Resource-Efficient Few-Shot Plant Disease Classification,” Dec. 15, 2025, arXiv: arXiv:2512.13428. doi: 10.48550/arXiv.2512.13428.

24. R. Ranjan, J. P. Singh, and A. K. Titoriya, “Few-shot learning for plant disease detection using DeepBDC,” Soft Comput, vol. 29, no. 17–18, pp. 5327–5340, Sep. 2025, doi: 10.1007/s00500-025-10884-6.
25. X. Jiang et al., “PlantCaFo: An efficient few-shot plant disease recognition method based on foundation models,” Plant Phenomics, vol. 7, no. 1, p. 100024, Mar. 2025, doi: 10.1016/j.plaphe.2025.100024.
26. “Evaluation of Different Few-Shot Learning Methods in the Plant Disease Classification Domain.” Accessed: Apr. 01, 2026. [Online]. Available: https://www.mdpi.com/2079-7737/14/1/99
27. S. Maladi, H. Akundi, P. Vanamada, and R. Bellamkonda, “Effective Plant Disease Identification Using Few-Shot Learning,” vol. 11, no. 12, 2024.
28. H. Lin, R. Tse, S.-K. Tang, Z. Qiang, and G. Pau, “Few-Shot Learning for Plant-Disease Recognition in the Frequency Domain,” Plants, vol. 11, no. 21, p. 2814, Jan. 2022, doi: 10.3390/plants11212814.
29. “Frontiers | Few-shot learning approach with multi-scale feature fusion and attention for plant disease recognition.” Accessed: Apr. 01, 2026. [Online]. Available: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.907916/full