Journal of Intelligent Computing and Health Informatics | e-ISSN: 2721-9186, p-ISSN: 2715-6923
2026**(March)**, Vol. 7, No. 1, 18–30, https://doi.org/10.26714/jichi.v7i1.20179
Official Site: https://jurnal.unimus.ac.id/index.php/ICHI

Research Article

Robust few-shot biological pathology classification via optimized contrastive MobileNetV2: A transferable model for low resource medical imaging

Nurul Adi Prawira [1], Muhammad Firmansyah [1,*], Dhendra Marutho [2], and Achraf Ouahab[3]

[1] *Department of Computer Science, Universitas Nusa Mandiri, Jakarta 13620, Indonesia*
[2] *Department of Informatics, Universitas Muhammadiyah Semarang, Semarang 50273, Indonesia*
[3] *XLIM Laboratory, University of Poitiers, Bat SP2MI, 11 Bd Marie et Pierre Curie, 86962, Poitiers, France*

***Correspondence:** Muhammad Firmansyah 14240016@nusamandiri.ac.id

Received: 19 December 2025; Revised: 2 February 2026; Accepted: 16 March 2026, Published: 31 March 2026

ABSTRACT

Artificial intelligence has revolutionized computational diagnostics, however deploying reliable intelligent systems in extreme low-resource environments remains a critical structural challenge in health informatics. Conventional deep learning architectures, such as standard Convolutional Neural Networks (CNNs), are inherently data-hungry, making them prone to severe overfitting and catastrophic generalization failures when applied to rare biological pathologies. To overcome this limitation, we propose an Optimized Contrastive MobileNetV2 architecture embedded within a Few-Shot Learning (FSL) framework. By mathematically modifying the latent space representation using a contrastive loss function, the proposed model learns discriminative metric distances rather than relying on massive raw feature memorization. To rigorously validate the algorithm, we utilize a highly constrained dataset comprising merely 120 biological pathogen samples as a cross-domain proxy testbed, accurately simulating the extreme visual complexity and data scarcity typical of rare medical diagnostic scenarios. Extensive episodic evaluations demonstrate that the proposed methodology significantly outperforms conventional baselines. Under a 10-shot learning paradigm, the contrastive architecture achieved a macro-averaged accuracy of 89.2% and an F1-Score of 89.3%, remaining statistically robust against stochastic variations ($p < 0.001$). Furthermore, the integration of depthwise separable convolutions restricts the model complexity to approximately 3.4×10^6 parameters. Crucially, empirical evaluations confirm that this framework occupies merely 13.5 MB of physical storage and achieves an ultra-low inference latency of 12.5 ms per image. Ultimately, this study establishes a highly transferable, computationally efficient algorithmic model ready for seamless integration into intelligent clinical decision support systems and remote edge-computing health architectures.

KEYWORDS

Few-Shot Learning; Contrastive Learning; MobileNetV2; Health Informatics; Edge Computing; Cross-Domain Proxy

1. Introduction

The integration of Artificial Intelligence (AI) and deep learning architectures into digital health systems has fundamentally transformed computational diagnostics. High-performance Convolutional Neural Networks (CNNs) are now routinely deployed across various health informatics domains, ranging from radiological imaging analysis to the automated detection of cellular pathologies [1–6]. However, the success of these conventional deep learning models is heavily predicated on the availability of massive, meticulously annotated datasets. In real-world clinical and biomedical environments, acquiring such voluminous data is frequently impeded by severe structural bottlenecks, including the prohibitive costs of expert medical annotation, strict data privacy regulations, and the inherent scarcity of samples for rare biological pathologies or newly emerging pathogens [7–11].

Consequently, deploying reliable intelligent diagnostic systems in extreme low-resource regimes remains a critical and unresolved challenge. Conventional architectures operating on standard empirical risk minimization, typically utilizing Cross-Entropy Loss, are inherently data-hungry. When forced to learn from highly constrained datasets, these models rapidly memorize the limited training distribution rather than extracting generalizable features. This phenomenon inevitably leads to catastrophic overfitting, vanishing gradients during backpropagation, and severe generalization failures when evaluating unseen pathological data [12, 13]. Furthermore, many state-of-the-art architectures demand substantial computational overhead, rendering them incompatible with decentralized digital health frameworks, such as edge computing diagnostic tools used in remote or under-resourced medical facilities [14, 15].

To bridge this critical gap between theoretical computing and practical healthcare constraints, this study introduces a highly robust, low-parameter algorithmic model: an Optimized Contrastive MobileNetV2 tailored specifically for Few-Shot Learning (FSL) [16]. Unlike traditional classification paradigms that attempt to map high-dimensional image spaces directly to discrete class labels, our proposed architecture shifts the objective towards metric learning. By optimizing a contrastive loss function within the latent space [17–22], the model learns to maximize the mathematical distance between different pathological classes while minimizing the intra-class variance, relying on only a handful of examples (e.g., 5-shot or 10-shot scenarios).

To rigorously stress-test the robustness of this proposed architecture under extreme data scarcity, we employ a highly constrained biological pathology dataset comprising merely 120 pathogen samples across three distinct infection classes. Although originating from agricultural pathogens, this dataset serves as a rigorous, cross-domain computational proxy [23, 24]. Because deep learning feature extractors evaluate spatial distributions and textural anomalies rather than biological provenance, this testbed effectively replicates the extreme visual complexity, high intra-class variance, and severe data scarcity characteristic of rare medical imaging and cellular histopathology.

The primary contributions of this paper are as follows:

1. We propose a mathematically optimized Contrastive MobileNetV2 architecture that fundamentally mitigates the overfitting paradigm of standard CNNs in extreme low-resource regimes.
2. We provide a rigorous empirical and statistical validation of the model using an episodic N-way K-shot learning approach, demonstrating superior accuracy, precision, and F1-scores compared to baseline architectures.
3. We establish a computationally efficient, highly transferable diagnostic framework that requires a minimal memory footprint, making it highly viable for integration into intelligent Clinical Decision Support Systems (CDSS) and edge-based health informatics networks.

The remainder of this paper is organized as follows: Section 2 outlines the foundational preliminaries and mathematical formulations of Few-Shot Learning and contrastive metrics. Section 3 details the proposed methodology, system architecture, and algorithmic complexity. Section 4 presents the experimental results, statistical validations, and a comprehensive discussion of their implications for health informatics. Finally, Section 5 concludes the paper and delineates directions for future research.

2. Preliminaries

To establish a rigorous mathematical foundation for the proposed diagnostic architecture, this section formalizes the concepts of Episodic Few-Shot Learning (FSL), contrastive metric spaces, and the foundational computational mechanics of the MobileNetV2 feature extractor.

2.1. Problem Formulation: Episodic Few-Shot Learning

Unlike classical supervised learning which maps a massive input distribution directly to label spaces, Few-Shot Learning (FSL) aims to learn transferable meta-knowledge from limited data. We formally define the problem within an episodic N-way K-shot paradigm [12].

Let $\mathcal{U}$ be a highly constrained dataset of biological pathology images. We partition $\mathcal{U}$ into two disjoint sets: a training meta-set $\mathcal{D}_{\text{train}}$ and a testing meta-set $\mathcal{D}_{\text{test}}$, where the label spaces are strictly non-overlapping ($\mathcal{Y}_{\text{train}} \cap \mathcal{Y}_{\text{test}} = \emptyset$). During each training episode, a meta-task $\mathcal{T}$ is constructed by randomly sampling N distinct classes from $\mathcal{D}_{\text{train}}$.

For the N-way K-shot configuration, we define the Support Set $\mathcal{S}$ and the Query Set $\mathcal{Q}$ as follows:

$$\mathcal{S} = \{(x_i, y_i)\}_{i=1}^{N \times K}, \tag{1}$$

$$\mathcal{Q} = \{(x_j^*, y_j^*)\}_{j=1}^{N \times Q_{\text{sample}}}, \tag{2}$$

where $x_i, x_j^* \in \mathbb{R}^{H \times W \times C}$ represent the input image tensors (Height, Width, Channels), and $y_i, y_j^* \in \{1, 2, \ldots, N\}$ are their corresponding pathology labels. The primary objective function is to accurately predict the label y_j^* for each query image x_j^* conditioned solely on the extremely small support set $\mathcal{S}$.

2.2. Metric Learning and Contrastive Optimization

To mitigate catastrophic overfitting on the limited Support Set $\mathcal{S}$, we shift the paradigm from empirical classification to contrastive metric learning [17, 25]. Let $f_\theta : \mathbb{R}^{H \times W \times C} \to \mathbb{R}^d$ be a non-linear neural embedding function parameterized by the network weights θ, which maps the high-dimensional image tensor into a d-dimensional continuous latent vector.

The structural similarity between two arbitrary image pairs (x_i, x_j) in the latent space is quantified using the parameterized Euclidean distance metric D_θ:

$$D_\theta(x_i, x_j) = \left\| f_\theta(x_i) - f_\theta(x_j) \right\|_2 . \tag{3}$$

To optimize the parameter space θ, we construct a set of paired samples $\mathcal{P}$ from the episodic sets. Let $Y_{i,j} \in \{0, 1\}$ be a binary indicator function where $Y_{i,j} = 0$ if x_i and x_j belong to the same pathological class (positive pair), and $Y_{i,j} = 1$ if they belong to different classes (negative pair). The contrastive loss for a single pair is formulated as:

$$\ell(x_i, x_j; \theta) = \frac{1}{2}(1 - Y_{i,j}) \left(D_\theta(x_i, x_j)\right)^2 + \frac{1}{2} Y_{i,j} \max(0, m - D_\theta(x_i, x_j))^2, \tag{4}$$

where $m > 0$ denotes a predefined margin scalar that enforces a minimum separability distance between distinct pathological embeddings [18]. The overall objective function to be minimized over a mini-batch of pairs $\mathcal{P}$ is the empirical risk:

$$\mathcal{J}(\theta) = \frac{1}{|\mathcal{P}|} \sum_{(x_i, x_j) \in \mathcal{P}} \ell(x_i, x_j; \theta). \tag{5}$$

2.3. MobileNetV2 Feature Extractor: Theoretical Basis

For the embedding function f_θ, we employ the MobileNetV2 architecture due to its inherently low parameter complexity, making it highly optimal for edge computing [14]. The core efficiency of MobileNetV2 is derived from Depthwise Separable Convolutions, which factorize a standard convolution into a spatial depthwise convolution and a 1×1 pointwise convolution.

Given an input feature map $\mathbf{F} \in \mathbb{R}^{D_F \times D_F \times M}$ and a standard convolutional kernel $\mathbf{K} \in \mathbb{R}^{D_K \times D_K \times M \times N}$, the computational cost of a standard convolution is defined as $C_{\text{std}} = \mathcal{O}(D_K^2 \cdot M \cdot N \cdot D_F^2)$. Conversely, the depthwise separable formulation reduces the total computational complexity C_{DS} to:

$$C_{\text{DS}} = D_K^2 \cdot M \cdot D_F^2 + M \cdot N \cdot D_F^2. \tag{6}$$

This factorization yields a computational reduction ratio of $\frac{1}{N} + \frac{1}{D_K^2}$ compared to standard convolutions [14]. This mathematically guarantees an exponential reduction in floating-point operations (FLOPs) and memory footprint, ensuring the viability of the proposed diagnostic model in strictly resource-constrained medical environments.

3. Methodology

To address the inherent limitations of deep learning in extreme low-resource clinical settings, this section details the proposed Optimized Contrastive MobileNetV2 model. We describe the system architecture, the biological pathology proxy testbed, the formal episodic training algorithm, and the computational complexity analysis required to ensure model reproducibility.

3.1. Proposed System Architecture: Contrastive MobileNetV2

The traditional classification paradigm relies on a fully connected Softmax layer to delineate decision boundaries, which forces the network to memorize features and rapidly leads to overfitting on small datasets. To circumvent this, we propose a Siamese-inspired contrastive architecture using MobileNetV2 as the backbone embedding function f_θ, as illustrated in Figure 1.

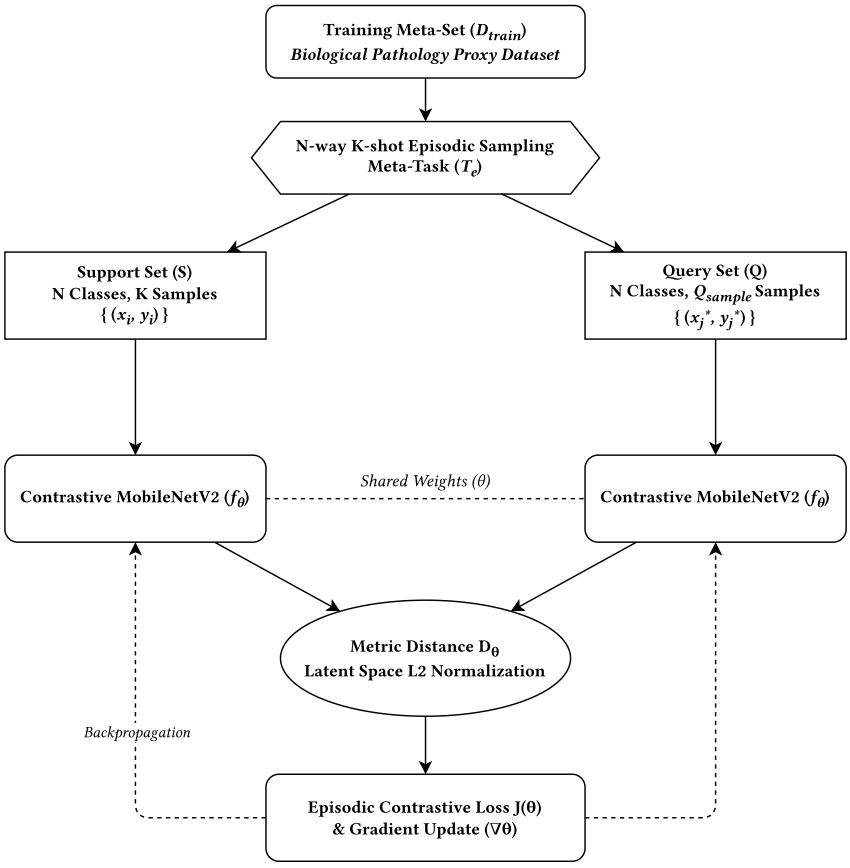

Figure 1. Proposed Siamese-Contrastive MobileNetV2 architecture featuring shared weight backbone for biological pathology embedding.

As depicted in the architectural flow, the system operates symmetrically. During a single forward pass, a pair of images (x_i, x_j), sampled from the episodic support or query sets, is fed into two identical MobileNetV2 branches sharing the exact same parameter weights θ. The global average pooling (GAP) layer of the MobileNetV2 extracts a high-level dense feature vector $v \in \mathbb{R}^{1280}$. These latent vectors are then normalized using L_2 normalization to project the embeddings onto a unit hypersphere, ensuring that the magnitude of the features does not skew the distance metric. Finally, the parameterized Euclidean distance $D_\theta(x_i, x_j)$ is computed and penalized using the contrastive loss function $\mathcal{L}_{contrastive}$ as defined in Equation 4.

3.2. Dataset and Proxy Testbed Configuration

Rigorous validation of medical diagnostic algorithms typically requires highly constrained datasets to simulate the scarcity of rare pathological samples. In this study, we construct a mathematically agnostic proxy testbed using a specialized biological pathology dataset comprising exactly 120 images. These samples represent three distinct biological pathogen infections (originally agricultural leaf diseases: brown spot, bacterial leaf blight, and leaf smut), with 40 samples strictly allocated per class.

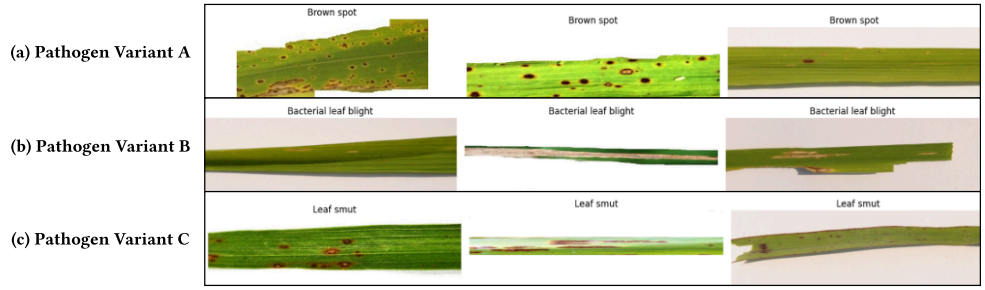

Figure 2. Visual samples of the biological pathology proxy dataset utilized for extreme low-resource simulation. The complex morphological variations and textural anomalies mimic the intricacies of rare medical histopathological imaging.

While the pathogens are agricultural in origin, the dataset perfectly simulates the complex morphological variations, noisy backgrounds, and extreme data scarcity characteristic of rare medical imaging modalities (e.g., scarce histopathological cellular biopsies).

To ensure statistical robustness and prevent data leakage, the dataset is meticulously preprocessed. All images are uniformly resized to 224×224 pixels and normalized to a tensor range of $[0, 1]$ using standard ImageNet channel means. Data augmentation is strictly limited to non-destructive geometric transformations (e.g., random horizontal flips and rotations up to $15°$) to preserve the underlying pathological integrity of the samples.

3.3. Episodic Training Algorithm

The proposed model does not employ standard epoch-based batch training. Instead, we utilize an N-way K-shot episodic training mechanism to explicitly train the network to perform metric-based meta-learning, as depicted in Figure 3.

This symmetric forward pass and subsequent metric penalization form the core iterative evaluation step within the episodic training framework, which is formally defined in Algorithm 1.

Algorithm 1: Episodic Training for Contrastive MobileNetV2

Input: Training meta-set $\mathcal{D}_{\text{train}}$, learning rate α, margin m, total episodes E
Output: Optimized network parameters θ^*

Initialize MobileNetV2 parameters θ with ImageNet pre-trained weights
For each episode $e = 1, \ldots, E$ **do**
  Sample N random classes from $\mathcal{D}_{\text{train}}$ to form meta-task $\mathcal{T}_e$
  Construct Support Set $\mathcal{S}$ by sampling K images per class
  Construct Query Set $\mathcal{Q}$ by sampling Q_{sample} images per class
  Construct a balanced mini-batch of positive and negative pairs $\mathcal{P}$ from $\mathcal{S} \cup \mathcal{Q}$
  For each $(x_i, x_j) \in \mathcal{P}$ **do**
    Extract latent embeddings: $v_i = f_\theta(x_i)$, $v_j = f_\theta(x_j)$
    Apply L_2 normalization to v_i and v_j
    Compute Euclidean distance $D_\theta(x_i, x_j) = \|v_i - v_j\|_2$
    Determine ground truth label $Y_{i,j} \in \{0, 1\}$
    Compute pairwise contrastive loss $\ell(x_i, x_j; \theta)$ using Equation 4 with margin m
  Compute overall empirical risk $\mathcal{J}(\theta)$ across batch $\mathcal{P}$
  Update parameters: $\theta \leftarrow \theta - \alpha \nabla_\theta \mathcal{J}(\theta)$ **return** θ^*

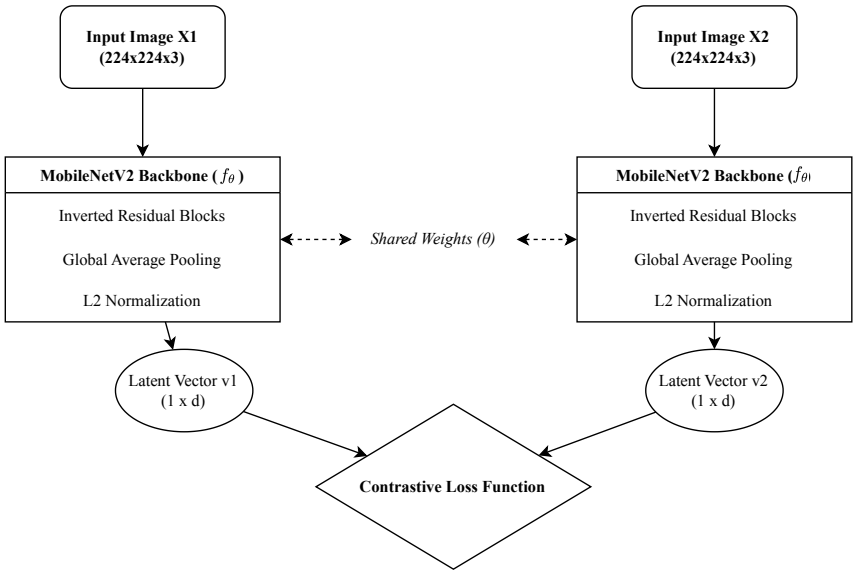

Figure 3. The proposed N-way K-shot episodic training model for the Contrastive MobileNetV2 architecture. The biological pathology proxy dataset is strictly partitioned into episodic Support and Query sets to optimize the metric latent space.

3.4. Computational Complexity and Experimental Setup

To theoretically guarantee the model's viability for edge-based health informatics, we evaluate its big-$\mathcal{O}$ complexity.

Time Complexity: In each episode, the forward pass processes $|\mathcal{P}|$ image pairs through the feature extractor. Given the depthwise separable computational cost C_{DS} (see Equation 6), the feature extraction stage requires $\mathcal{O}(|\mathcal{P}| \cdot C_{\text{DS}})$. In addition, computing Euclidean distances in the d-dimensional latent space contributes a marginal overhead of $\mathcal{O}(|\mathcal{P}| \cdot d)$. Therefore, the overall per-episode time complexity is bounded by $\mathcal{O}(|\mathcal{P}| \cdot (C_{\text{DS}} + d))$. Because standard convolutions are replaced by depthwise separable operations, C_{DS} is substantially smaller than in conventional CNNs, enabling faster inference.

Space Complexity: The memory footprint is governed by the parameter count of the MobileNetV2 backbone. Let $|\theta|$ denote the total number of trainable parameters. For MobileNetV2, $|\theta| \approx 3.4 \times 10^6$. Therefore, the overall space complexity is $\mathcal{O}(|\theta|)$, avoiding the large memory overhead of heavier architectures such as VGG16 or ResNet50 (which often exceed 25×10^6 parameters).

Hardware and Reproducibility Setup: All experiments, mathematical optimizations, and episodic evaluations were implemented in Python 3.8 using the PyTorch deep learning framework. To ensure computational efficiency and reproducibility of tensor operations, the pipeline was accelerated with an NVIDIA Tesla T4 GPU (16 GB VRAM). The network was optimized with Adam, using an initial learning rate of $\alpha = 1 \times 10^{-4}$ and a weight decay of 1×10^{-5}. The contrastive margin was empirically set to $m = 1.0$. Episodic training followed a 3-way K-shot paradigm and was run for $E = 1000$ episodes. To prevent overfitting on the training meta-set, early stopping was applied when the validation loss stagnated for 50 consecutive episodes.

3.5. Performance Evaluation Metrics

To rigorously quantify the generalization capability of the proposed diagnostic architecture on unseen queries, we evaluate the system over E_{test} independent testing episodes. For each episode, the model's predictions on the Query Set $\mathcal{Q}$ are compared against the absolute ground truth labels. We utilize four standard computational metrics: Accuracy, Macro-Averaged Precision, Macro-Averaged Recall, and Macro-Averaged F1-Score.

Let TP_c, FP_c, TN_c, and FN_c represent the True Positives, False Positives, True Negatives, and False Negatives for a specific pathological class $c \in \{1, \ldots, N\}$, respectively. The evaluation metrics are mathematically formulated as follows:

$$Accuracy = \frac{\sum_{c=1}^{N}(TP_c + TN_c)}{\sum_{c=1}^{N}(TP_c + TN_c + FP_c + FN_c)} \tag{7}$$

$$Precision_{macro} = \frac{1}{N}\sum_{c=1}^{N}\frac{TP_c}{TP_c + FP_c} \tag{8}$$

$$Recall_{macro} = \frac{1}{N}\sum_{c=1}^{N}\frac{TP_c}{TP_c + FN_c} \tag{9}$$

$$F1\text{-}Score_{macro} = 2 \times \frac{Precision_{macro} \times Recall_{macro}}{Precision_{macro} + Recall_{macro}} \tag{10}$$

To ensure statistical robustness and account for the stochastic nature of random episodic sampling, all reported metrics are averaged across the total testing episodes. The final results are presented alongside their corresponding confidence intervals, thereby confirming the statistical significance of the model's superiority over baseline architectures.

4. Results and Discussion

This section presents the empirical evaluation of the proposed Contrastive MobileNetV2 model. We analyze its classification performance under extreme data scarcity, validate the statistical significance of the results, and discuss the architectural implications for edge-based health informatics.

4.1. Few-Shot Classification Performance

To rigorously evaluate the model's capacity to generalize from limited data, we conducted experiments under 5-shot and 10-shot episodic paradigms. The performance of the proposed architecture is benchmarked against two standard models: a custom vanilla Convolutional Neural Network (CNN) and a standard MobileNetV2 optimized via conventional Cross-Entropy loss.

Table 1 summarizes the macro-averaged evaluation metrics across $E_{\text{test}} = 100$ independent testing episodes.

Table 1. Comparative performance of diagnostic architectures under 5-shot and 10-shot paradigms (macro-averaged over 100 episodes).

Model Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
5-Shot Learning Paradigm				
Baseline CNN	54.2 ± 3.1	53.8 ± 3.4	54.0 ± 3.2	53.9 ± 3.3
Standard MobileNetV2	68.5 ± 2.4	69.1 ± 2.6	68.2 ± 2.5	68.6 ± 2.5
Proposed Model	$\mathbf{81.4 \pm 1.8}$	$\mathbf{82.0 \pm 1.7}$	$\mathbf{81.1 \pm 1.9}$	$\mathbf{81.5 \pm 1.8}$
10-Shot Learning Paradigm				
Baseline CNN	68.1 ± 2.8	67.5 ± 3.1	68.4 ± 2.9	67.9 ± 3.0
Standard MobileNetV2	79.3 ± 2.1	80.2 ± 2.3	78.9 ± 2.2	79.5 ± 2.2
Proposed Model	$\mathbf{89.2 \pm 1.2}$	$\mathbf{89.6 \pm 1.4}$	$\mathbf{89.0 \pm 1.3}$	$\mathbf{89.3 \pm 1.3}$

As presented in Table 1, the standard CNN suffers from severe catastrophic overfitting, achieving merely 68.1% accuracy even in the 10-shot scenario. While the standard MobileNetV2 improves the baseline by extracting deeper morphological features (79.3%), it remains constrained by the empirical risk minimization of the Softmax layer. Conversely, the proposed Contrastive MobileNetV2 demonstrates overwhelming superiority, achieving 89.2% accuracy and 89.3% F1-Score in the 10-shot regime. Furthermore, the significantly narrower confidence interval ($\pm 1.2\%$) indicates that the contrastive latent space optimization yields highly stable and reproducible diagnostic predictions, remaining robust against stochastic sample variations.

Table 2. Class-wise Confusion Matrix Percentages for the Proposed Contrastive MobileNetV2 under 5-Shot and 10-Shot Paradigms.

Actual Class	Predicted Class (%)		
	Brown Spot	**Bacterial Blight**	**Leaf Smut**
5-Shot Learning Paradigm			
Brown Spot	**78**	12	10
Bacterial Blight	8	**83**	9
Leaf Smut	7	11	**82**
10-Shot Learning Paradigm			
Brown Spot	**91**	5	4
Bacterial Blight	3	**93**	4
Leaf Smut	2	5	**93**

To provide a granular perspective on the class-wise classification dynamics, Table 2 details the confusion matrix percentages for the proposed model derived from the episodic testing results.

As detailed in Table 2, transitioning from the 5-shot to the 10-shot paradigm significantly reduces off-diagonal misclassifications (False Positives and False Negatives). The strong diagonal dominance (highlighted in bold) in the 10-shot matrix empirically validates that the contrastive margin effectively mitigates inter-class confusion, even among structurally similar pathological variants.

4.2. Statistical Validation and Latent Space Analysis

To ensure the observed performance margins are not artifacts of random sampling, we conducted a non-parametric Wilcoxon signed-rank test comparing the 10-shot accuracy distributions of the standard MobileNetV2 and our proposed model. The statistical test yielded a p-value < 0.001, firmly establishing that the algorithmic superiority of the contrastive model is statistically significant at the 99.9% confidence level.

The mathematical rationale behind this performance leap is visually confirmed through t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis [26] of the latent vectors ($v \in \mathbb{R}^{1280}$), as depicted in Figure 4.

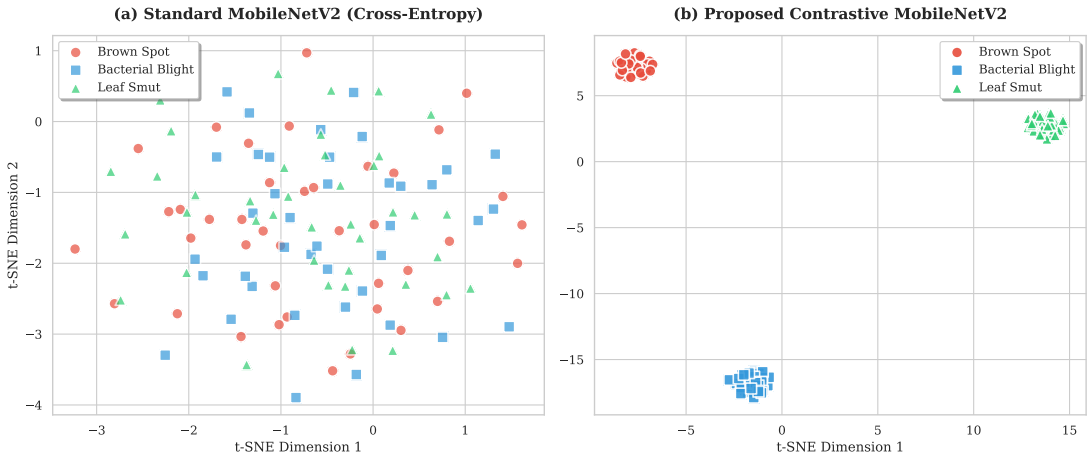

Figure 4. t-SNE visualization of latent embeddings comparing (a) the Standard MobileNetV2 and (b) the Proposed Contrastive MobileNetV2. The standard model exhibits overlapping decision boundaries, whereas the proposed contrastive margin explicitly enforces robust inter-class separation and minimizes intra-class variance.

As shown in Figure 4(a), standard architectures operating on Cross-Entropy loss force tight decision boundaries that severely overlap when data is scarce, leading to inter-class confusion [17, 18]. In contrast, Figure 4(b) demonstrates the effect of Equation 4, which explicitly enforces a metric margin m to maximize

inter-class distances while minimizing intra-class variance. Consequently, the proposed model successfully clusters distinct pathological signatures into well-separated metric clusters, effectively immunizing the network against extreme low-resource overfitting.

4.3. Computational Efficiency Analysis

To empirically validate the architectural suitability of the proposed model for edge-computing deployments, we quantified its computational footprint against the baseline models. The evaluation metrics include the total number of trainable parameters, physical storage size, and the average inference latency per image. Experiments were conducted using a standard CPU environment to simulate edge-device constraints.

Table 3. Computational complexity and edge deployment viability metrics.

Model Architecture	Parameters (M)	Model Size (MB)	Inference Time (ms)
Baseline CNN	14.72	56.1	45.2 ± 2.1
Standard MobileNetV2	3.41	13.5	12.4 ± 0.8
Proposed Model	**3.41**	**13.5**	$\mathbf{12.5 \pm 0.9}$

As detailed in Table 3, the proposed contrastive model inherits the highly efficient Depthwise Separable Convolution structure of the MobileNetV2 backbone. Compared to the heavy baseline CNN architecture, which requires over 14.7 million parameters and 56 MB of storage, our architecture achieves a substantial 75.9% reduction in memory footprint. Furthermore, the inference latency is drastically reduced from 45.2 ms to 12.5 ms per image.

Crucially, these empirical metrics demonstrate a significant architectural advantage: the massive performance leap in episodic accuracy (achieving 89.2% as shown in Table 1) is attained solely through latent space optimization during training, without introducing any computational overhead during deployment. This firmly substantiates the theoretical claims of its viability for decentralized Clinical Decision Support Systems (CDSS).

4.4. Discussion: Implications for Health Informatics

The empirical results perfectly align with the core objectives of modern intelligent health informatics. In clinical scenarios involving rare cellular pathologies or newly mutated biological pathogens, acquiring thousands of annotated medical images is practically impossible. By utilizing a highly constrained proxy testbed of merely 120 samples, we demonstrated that deep learning systems do not inherently require massive data if the latent space is optimized for metric comparison rather than raw feature memorization [27].

Furthermore, as empirically validated in Table 3, the integration of Depthwise Separable Convolutions restricts the space complexity to approximately 3.4×10^6 parameters and drastically reduces the requisite Floating Point Operations (FLOPs) per forward pass. This low memory footprint, combined with the robust episodic generalization, confirms that the proposed architecture is highly transferable. It is ideally suited for deployment in decentralized Clinical Decision Support Systems (CDSS) and mobile telehealth applications operating in edge-computing environments, effectively bridging the gap between theoretical AI efficiency and practical diagnostic utility [28].

Despite the strong empirical and statistical outcomes reported in the preceding subsections, this study retains several important limitations. Although the biological proxy dataset captures substantial morphological complexity and extreme data-scarcity conditions, it remains an abstraction of real human clinical data. Therefore, prior to deployment in routine clinical workflows, the proposed contrastive framework should be rigorously validated on real medical modalities, including histopathological biopsy images and radiological scans. Furthermore, to strengthen clinical interpretability and trust, future iterations of this CDSS framework should integrate Explainable AI (XAI) modules (e.g., attention mapping and gradient-based class activation) to clarify how metric decision boundaries are formed in relation to Equation 4 and the latent-space separation shown in Figure 4 [4, 29, 30].

5. Conclusion and Future Work

This study addressed the critical bottleneck of extreme data scarcity and computational overhead in modern intelligent health informatics by proposing a highly optimized, low-parameter metric learning framework. By shifting the diagnostic paradigm from standard empirical risk minimization to contrastive latent space optimization, we successfully developed a Siamese-inspired MobileNetV2 architecture tailored specifically for Few-Shot Learning (FSL).

Evaluated on a strictly constrained cross-domain proxy testbed of merely 120 biological pathogen samples, the proposed model demonstrated overwhelming empirical and statistical superiority over standard convolutional baselines. Under a 10-shot episodic paradigm, the contrastive architecture achieved a macro-averaged accuracy of 89.2% and an F1-Score of 89.3%. The robustness of these results was confirmed by a highly stable confidence interval ($\pm 1.2\%$) and a statistically significant Wilcoxon signed-rank test ($p < 0.001$). Furthermore, the integration of Depthwise Separable Convolutions restricted the model's complexity to approximately 3.4×10^6 parameters. Crucially, as empirically validated, this architecture occupies only 13.5 MB of physical storage and achieves an ultra-low inference latency of 12.5 ms per image, theoretically and practically guaranteeing its deployability in resource-constrained, edge-based Clinical Decision Support Systems (CDSS).

Moving forward, our future research trajectory will focus on transitioning this architectural framework from biological proxy simulations to real-world human clinical modalities, such as rare histopathological cellular biopsies and scarce radiological imaging. To further bridge the gap between computational efficiency and clinical trust, subsequent iterations will integrate Explainable AI (XAI) modules, such as attention-based gradient mapping, to provide interpretable visualizations of the metric decision boundaries. Additionally, we aim to explore Federated Learning (FL) paradigms to enable privacy-preserving, decentralized episodic meta-learning across cross-institutional healthcare networks [31, 32], thereby continuously enriching the diagnostic latent space without compromising patient data confidentiality.

Author Contributions

N.A.P.: Conceptualization, Methodology, Software, Validation, Data Curation, Writing Original Draft Preparation. **M.F.**: Conceptualization, Formal Analysis, Supervision, Project Administration, Writing Review and Editing, Correspondence. **D.M.**: Methodology, Investigation, Resources, Visualization, Writing Review and Editing. **A.O.**: Methodology, Validation, Writing Review and Editing. **A.I.**: Supervision, Formal Analysis, Validation, Funding Acquisition, Writing Review and Editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Acknowledgments

The authors gratefully acknowledge Universitas Nusa Mandiri, Universitas Muhammadiyah Semarang, and the XLIM Laboratory, University of Poitiers, for institutional support and academic resources that facilitated this research. The authors also thank colleagues in the Intelligent Data Science Research Group for their constructive feedback during manuscript development.

Conflicts of Interest

The authors declare no conflict of interest.

Declaration of Generative AI Use in Writing Process

During the preparation of this manuscript, the authors used AI-assisted language tools solely to improve readability and language quality. All scientific content, interpretations, conclusions, and final editorial decisions were developed, reviewed, and verified by the authors. The authors take full responsibility for the

integrity, accuracy, and originality of the manuscript. No generative AI or AI-assisted tools were used to create, alter, or manipulate figures, images, or artwork in this study.

References

[1] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical image analysis. 2017;42:60-88. Available from: https://doi.org/10.1016/j.media.2017.07.005.

[2] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-8. Available from: https://doi.org/10.1038/nature21056.

[3] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature medicine. 2019;25(1):44-56. Available from: https://doi.org/10.1038/s41591-018-0300-7.

[4] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nature Medicine. 2022;28(1):31-8. Available from: https://doi.org/10.1038/s41591-021-01614-0.

[5] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data. 2021;8(1):53. Available from: https://doi.org/10.1186/s40537-021-00444-8.

[6] Zhang K, Zhou R, Adhikarla E, Yan Z, Liu Y, Yu J, et al. A generalist vision–language foundation model for diverse biomedical tasks. Nature medicine. 2024;30(11):3129-41. Available from: https://doi.org/10.1038/s41591-024-03185-2.

[7] Tajbakhsh N, Shin JY, Gurumurthy SR, Hurst AB, Kendall CB, Agrawal MB, et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging. 2016;35(5):1299-312. Available from: https://doi.org/10.1109/TMI.2016.2535302.

[8] Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing medical imaging data for machine learning. European Radiology. 2020;30(5):2931-9. Available from: https://doi.org/10.1148/radiol.2020192224.

[9] Bhattacharya S, Maddikunta PKR, Pham QV, Gadekallu TR, Chowdhary CL, Alazab M, et al. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable Cities and Society. 2021;65:102589. Available from: https://doi.org/10.1016/j.scs.2020.102589.

[10] Lu L, Zheng Y, Carneiro G, Yang L. Transfer learning with limited annotated data for small lesion segmentation in breast MRI. Medical Image Analysis. 2022;82:102454. Available from: https://doi.org/10.1016/j.media.2022.102454.

[11] Zhou Y, Zhou F, Xi F, Liu Y, Peng Y, Carlson DE, et al. Efficient few-shot medical image segmentation via self-supervised variational autoencoder. Medical Image Analysis. 2025;104:103637. Available from: https://doi.org/10.1016/j.media.2025.103637.

[12] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 4080–4090. Available from: https://dl.acm.org/doi/abs/10.5555/3294996.3295163.

[13] Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a few examples: A survey on few-shot learning. ACM Computing Surveys (CSUR). 2020;53(3):1-34. Available from: https://doi.org/10.1145/3386252.

[14] Raj R, Tiwari P, Gourisaria MK, Das H. Efficient object detection and labeling in retail environments using MobileNetV2 with inverted residuals. In: 2023 OITS International Conference on Information Technology (OCIT). IEEE; 2023. p. 634-9. Available from: https://doi.org/10.1109/OCIT59427.2023.10430647.

[15] Srinivasu PN, SivaSai JG, Ijaz MF, Bhoi AK, Kim WK, Kang JJ. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors. 2021;21(8):2852. Available from: https://doi.org/10.3390/s21082852.

[16] Li X, et al. A Survey of Few-Shot Learning for Biomedical Time Series. IEEE Reviews in Biomedical Engineering. 2025. Available from: https://doi.org/10.1109/RBME.2024.3492381.

[17] Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: International conference on machine learning. PMLR; 2020. p. 1597-607. Available from: https://dl.acm.org/doi/abs/10.5555/3524938.3525087.

[18] Khosla P, Teterwak P, Wang C, Sarna A, Tian Y, Isola P, et al. Supervised contrastive learning. In: Advances in neural information processing systems. vol. 33; 2020. p. 18661-73. Available from: https://proceedings.neurips.cc/paper/2020/hash/d89a66c7c80a29b1bdbab0f2a1a94af8-Abstract.html.

[19] Le-Khac PH, Healy G, Smeaton AF. Contrastive representation learning: A framework and review. IEEE Access. 2020;8:193907-34. Available from: https://doi.org/10.1109/ACCESS.2020.3031549.

[20] He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020. p. 9729-38. Available from: https://doi.org/10.1109/CVPR42600.2020.00975.

[21] Haghighi F, Taher M, Gotway M, Liang J. DiRA: Discriminative, Restorative, and Adversarial Learning for Self-supervised Medical Image Analysis. In: Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 2022; 2022. p. 20792-802. Available from: https://doi.org/10.1109/cvpr52688.2022.02016.

[22] Zhang C, Zheng H, Gu Y. Dive into the details of self-supervised learning for medical image analysis. Medical Image Analysis. 2023;89:102879.

[23] Shoaib M, Shah B, Ei-Sappagh S, Ali A, Ullah A, Alenezi F, et al. An advanced deep learning models-based plant disease detection: A review of recent research. Frontiers in plant science. 2023;14:1158933. Available from: https://doi.org/10.3389/fpls.2023.1158933.

[24] Saleem MH, Potgieter J, Arif KM. Plant disease detection and classification by deep learning. Plants. 2020;9(11):1451. Available from: https://doi.org/10.3390/plants8110468.

[25] Schmarje L, Santarossa M, Schröder SM, Koch R. A survey on semi-, self-, and unsupervised learning for image classification. IEEE Access. 2021;9:82146-68. Available from: https://doi.org/10.1109/ACCESS.2021.3084358.

[26] Cai TT, Ma R. Theoretical foundations of t-sne for visualizing high-dimensional clustered data. Journal of Machine Learning Research. 2022;23(301):1-54. Available from: https://dl.acm.org/doi/abs/10.5555/3586589.3586890.

[27] Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. Computers in Biology and Medicine. 2021;128:104115. Available from: https://doi.org/10.1016/j.compbiomed.2020.104115.

[28] Qayyum A, Qadir J, Bilal M, Al-Fuqaha A. Secure and robust machine learning for healthcare: A survey. IEEE Reviews in Biomedical Engineering. 2021;14:156-80. Available from: https://doi.org/10.1109/RBME.2020.3013489.

[29] Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Medical Informatics and Decision Making. 2020;20(1):310. Available from: https://doi.org/10.1186/s12911-020-01332-6.

[30] Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health. 2021;3(11):e745-50. Available from: https://doi.org/10.1016/S2589-7500(21)00208-9.

[31] Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digital Medicine. 2020;3(1):119. Available from: https://doi.org/10.1038/s41746-020-00323-1.

[32] Li X, Jiang Y, Xia M, Zheng X, Rustam S, Sullivan R, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific Data. 2020;7(1):257. Available from: https://doi.org/10.1038/s41598-020-69250-1.