Journal of Advanced College of Engineering and Management

# Malaria Parasite Classification in Thin Blood Smears Based on CNN with SpinalNet and GAN

[1]Bishwas Dangol, [2*]Sanjeeb Prasad Panday
*[1]MSc. in Data Science and Analytics Engineering, Pulchowk Campus, Lalitpur, Bagmati, Nepal*
*[2*]Associate Professor, Pulchowk Campus, Lalitpur, Bagmati, Nepal*
*Email: bishwasdangol1@gmail.com*
*Corresponding email: [*]sanjeeb@ioe.edu.np*
*DOI: 10.3126/jacem.v12i01.93926*

***Abstract***
This research presents a novel framework integrating Generative Adversarial Networks (GANs) with a parameter-efficient Convolutional Neural Network (CNN) to automate malaria parasite classification from thin blood smear microscopy images. A major bottleneck in automated malaria diagnosis is severe clinical data imbalance, where minority species (e.g., Plasmodium ovale, P. malariae) are heavily underrepresented, causing standard deep learning models to default to majority-class predictions. To address this, our methodology employs a Wasserstein GAN with Gradient Penalty (WGAN-GP) to synthesize high-fidelity microscopic images, effectively balancing the training distribution prior to classification. Following rigorous preprocessing---including Contrast Limited Adaptive Histogram Equalization (CLAHE) and aggressive data augmentation to mitigate staining variability---the balanced dataset is processed by a CNN for feature extraction. Crucially, traditional fully connected layers are replaced with a SpinalNet classification head. This architectural innovation processes extracted features sequentially, which drastically reduces the number of trainable dense parameters while acting as a structural regularizer against overfitting on the synthetic data. Evaluated on strictly real, hold-out clinical test sets, the proposed GAN-CNN-SpinalNet pipeline demonstrates a strong capability to identify rare parasite species without sacrificing performance on the P. falciparum majority class. Assessed across standard metrics, including accuracy and Macro F1-score, this approach successfully overcomes data scarcity to deliver a robust, and computationally efficient diagnostic tool for resource-limited settings.

***Keywords****—Malaria Diagnosis, Deep Learning, SpinalNet, CNN, Medical Image Analysis, GAN*

## 1. INTRODUCTION

Malaria continues to pose a significant global health burden, and its effective management critically depends on accurate and timely diagnosis. The current gold standard for diagnosis is the microscopic examination of thin stained blood smears. While effective, this manual process is inherently time-consuming and requires a high level of expertise

from trained microscopists. These challenges are severely exacerbated when attempting to identify rare malaria species, such as Plasmodium ovale or P. malariae. In real-world clinical settings, the vast majority of cases belong to P. falciparum, creating an extreme morphological data imbalance that makes comprehensive and accurate diagnosis difficult.

Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated considerable potential in automating the analysis of medical images. However, conventional CNNs applied to real-world malaria diagnosis face two primary limitations. First, they are highly susceptible to clinical data imbalance; when trained on heavily skewed datasets, standard CNNs tend to blindly bias toward the majority class, failing entirely to detect critical minority species. Second, modern CNN architectures often rely on traditional fully connected classification heads that contain millions of dense parameters. This architectural bloat not only demands high computational memory but also exacerbates the risk of overfitting when training on limited clinical data.

To overcome the critical challenge of minority class scarcity, this research leverages Generative Adversarial Networks (GANs) to augment the training environment. Specifically, a Wasserstein GAN with Gradient Penalty (WGAN-GP) is employed to synthesize high-fidelity microscopic images of the underrepresented parasite species. By injecting these biologically realistic synthetic images into the training pipeline, the dataset is effectively balanced. This prevents the neural network from taking statistical shortcuts, forcing it to actively learn the distinct morphological features of rare parasites.

Furthermore, to address the computational inefficiency and overfitting tendencies of standard CNN classifiers, this study integrates a SpinalNet architecture into the classification head. Unlike traditional multi-layer perceptrons that map all extracted features simultaneously to massive dense layers, SpinalNet breaks the CNN's extracted features into smaller, sequential subsets. This biologically inspired design drastically reduces the number of trainable parameters, acting as a powerful structural regularizer that prevents the model from overfitting on the GAN-generated data while maintaining high classification accuracy.

In summary, this paper proposes a robust, automated classification framework that integrates data-level balancing through WGAN-GP synthesis with an architecturally efficient Custom CNN and SpinalNet classifier. The objective is to deliver a reliable, highly accurate, and computationally lightweight tool capable of distinguishing between morphologically similar malaria species, offering a scalable diagnostic solution for resource-limited settings.

## 2. RELATED WORKS

Recent years have seen a proliferation of deep learning models tailored for parasite classification. The discussion below groups existing studies into three main areas: (1) CNN-based malaria classification, (2) GANs for medical image synthesis and dataset balancing, and (3) parameter-efficient classification heads and structural regularization.

### A. CNN-Based Malaria Classification

Significant research efforts have been dedicated to designing models that accurately classify infected cells according to parasite species and life-cycle stages. Early applications of deep learning in this domain relied on standard architectures like ResNet, VGG, and DenseNet. For example, the landmark study by Rajaraman et al. [1] extensively evaluated pre-trained CNNs as feature extractors for malaria detection. While they demonstrated that networks like ResNet-50 could achieve high binary accuracy (infected vs. uninfected), their work also highlighted that these deep architectures require massive amounts of balanced, annotated data to generalize effectively. Consequently, they consistently underperformed in multi-class scenarios involving rare parasite species (e.g., Plasmodium ovale) due to severe data scarcity in clinical samples [2].

To improve feature extraction without relying on massive models, lightweight CNN architectures have been proposed for mobile and point-of-care deployment. Modified versions of MobileNet and EfficientNet have been heavily utilized because their depthwise separable convolutions drastically reduce computational requirements. However, when applied to highly imbalanced microscopic datasets, these networks still suffer from a strong bias toward the majority class (typically P. falciparum), resulting in poor macro-level diagnostic metrics [3].

Researchers have attempted to solve this by integrating attention mechanisms within CNNs. The Soft Attention Parallel CNN (SPCNN) employs dual feature pathways with attention fusion to improve focus on infected regions, achieving high accuracy for binary tasks but lacking fine-grained species discrimination [4]. Similarly, a Stacked-LSTM model augmented with an attention mechanism achieved 99.12% accuracy for binary infection detection but did not easily scale to multi-class, multi-species classification [5]. The persistent failure of standard CNNs to effectively diagnose rare species highlights the limitations of purely algorithmic modifications when the underlying dataset is fundamentally skewed.

### B. Generative Adversarial Networks (GANs) for Dataset Balancing

To address the critical bottleneck of data scarcity and class imbalance, researchers have increasingly turned to data-level solutions, most notably Generative Adversarial Networks (GANs). GANs consist of two competing neural networks—a generator and a discriminator—that learn the underlying distribution of a dataset to synthesize highly realistic artificial images.

The application of GANs to overcome medical data scarcity was prominently established by Frid-Adar et al. [6], who proved that using synthetic images to augment imbalanced medical datasets significantly improves the sensitivity and specificity of CNN classifiers. Building on this in the context of malaria, Sengar et al. [7] successfully applied GANs to synthesize additional training samples for rare

Plasmodium vivax life cycle stages. By training a Vision Transformer (ViT) on this GAN-augmented dataset, they achieved 90.03% accuracy in classifying various parasite stages, proving that synthetic data can provide the necessary variance for deep learning models to generalize beyond majority classes.

Further advancing this concept, Tan and Liang [8] proposed a robust framework for multiclass malaria parasite recognition that explicitly paired GAN-based data augmentation with advanced transformer models. Their study demonstrated that synthesizing high-quality images of minority classes prior to training significantly mitigated the effects of clinical data imbalance, allowing the transformer's global self-attention mechanisms to accurately classify multiple species without defaulting to the majority class.

However, while coupling GANs with transformer-based architectures yields high accuracy, it introduces immense computational overhead. Standard GANs also suffer from training instability and "mode collapse." To counter this, the Wasserstein GAN with Gradient Penalty (WGAN-GP), introduced by Gulrajani et al. [9], replaces the standard Jensen-Shannon divergence with the Wasserstein distance metric and enforces a gradient penalty, ensuring smoother gradient flow and preventing mode collapse. While WGAN-GP has seen success in broader medical imaging, its specific application to balancing multi-species malaria datasets remains an underexplored but highly promising frontier.

### C. Parameter-Efficient Classification and SpinalNet

While GANs successfully solve the data imbalance problem, introducing thousands of synthetic images into a training pipeline often leads to overfitting, particularly in the massive, fully connected dense layers of traditional CNN classification heads. A standard multi-layer perceptron (MLP) head maps all extracted features to all hidden neurons simultaneously, resulting in millions of trainable parameters that demand high computational memory and are prone to memorizing synthetic artifacts rather than learning generalized biological features.

To mitigate this architectural bloat, researchers have explored structural regularization techniques. Recently, the SpinalNet [10] architecture has emerged as a highly efficient alternative to the standard MLP. Inspired by the biological somatosensory system of the human spine, SpinalNet divides the input features extracted by a CNN backbone into smaller, sequential blocks. Rather than passing all features at once, it feeds these smaller subsets iteratively into hidden layers, combining them with the outputs of previous layers.

This sequential, modular design drastically reduces the multiplicative mathematical operations required in the dense layers, often resulting in a 50% to 60% reduction in trainable parameters compared to a standard classification head. Empirical studies have

demonstrated that this parameter reduction acts as an inherent regularizer, naturally preventing overfitting while maintaining or even exceeding state-of-the-art accuracy.

Despite its proven efficiency in general computer vision tasks, the integration of SpinalNet with custom CNNs for multi-class microscopic parasite classification has not been widely documented. The combination of data-level balancing via WGAN-GP and architectural regularization via SpinalNet represents a novel, computationally lightweight, and holistic approach to solving the primary challenges of automated malaria diagnosis without the overhead of massive transformer models.

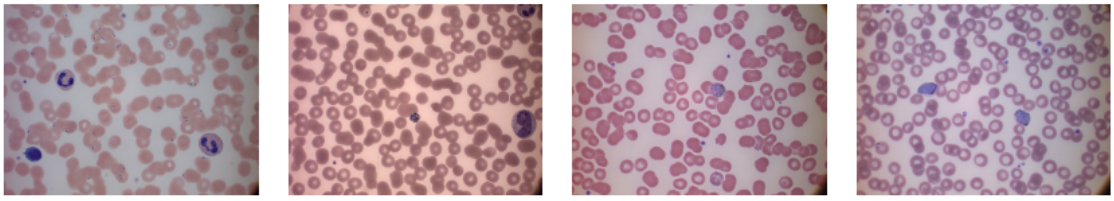
## 3. DATASET

Acquiring high-quality clinical data is a critical prerequisite for developing robust medical AI systems. Initial efforts to collect a localized clinical cohort involved consulting Teku Hospital (Kathmandu, Nepal), a central public health hub for infectious and tropical diseases. However, while patient diagnostic outcomes (positive or negative) are routinely recorded, the physical microscopic thin blood smear images are not systematically digitized or archived for retrospective computational analysis.

In the absence of a localized digital cohort, a composite dataset was assembled from multiple widely recognized, publicly available sources to ensure the model's robustness and global generalizability. This approach yielded a large volume of images with diverse staining variations, encompassing various Plasmodium species and life-cycle stages. The key public datasets utilized are summarized in Table 1.

While several of these repositories provide full-slide microscopy images with bounding box or polygon annotations originally intended for object detection, the individual infected cells (regions of interest) were extracted to construct a focused, single-cell dataset for the proposed classification pipeline.

Crucially, datasets such as the Malaria Parasite Image Database (MP-IDB [11]) provide detailed annotations for the four primary human-infecting species: Plasmodium falciparum, P. malariae, P. ovale, and P. vivax (illustrated in Figure 1). However, this composite data reflects the natural epidemiological prevalence of the disease, resulting in a severe, real-world class imbalance. P. falciparum instances heavily outnumber the minority classes, with species like P. ovale and P. malariae being exceedingly rare. This inherent clinical skew directly motivated the integration of WGAN-GP in the proposed methodology to artificially balance the dataset prior to CNN-SpinalNet classification.



(a) Falciparum (b) Malariae (c) Ovale (d) Vivax

**Figure 1**: Sample images showing different parasite species in thin blood smears from the MP-IDB dataset: (a) P. falciparum, (b) P. malariae, (c) P. ovale, and (d) P. vivax. Source: [11]

**Table 1**: Summary of major publicly available malaria parasite image datasets

| Dataset Name | Abbreviation | Total Images | Key Features & Classes |
|---|---|---|---|
| National Library of Medicine Dataset | NIH-NLM | 800 | Images from 160 patients. Detailed polygonal annotations for precise segmentation. |
| Istituto Mario Negri Malaria Dataset | IML | 346 | Annotations include bounding boxes and class labels. Focus on life cycle stage classification. |
| Malaria Parasite Image Database | MP-IDB | 1,212* | Contains four Plasmodium species: P. falciparum, P. malariae, P. ovale, P. vivax. Each species has four life stages: Ring, Trophozoite, Schizont, Gametocyte. |
| Malaria Bounding Box Dataset | MBB | 1,328 | Large volume of images. Annotations include bounding boxes and class labels for life cycle stages. |
| National Library of Medicine Dataset | NIH-NLM | 800 | Images from 160 patients. Detailed polygonal annotations for precise segmentation. |

## 4. METHODOLOGY

For the robust multiclass classification of malaria parasites from thin blood smear microscopy, a comprehensive computational framework was developed to overcome two fundamental bottlenecks in automated hematology: extreme epidemiological data imbalance and parameter-induced overfitting. To achieve high diagnostic fidelity across both rare and prevalent species, the proposed methodology systematically integrates morphological image preprocessing, generative dataset synthesis, hierarchical feature extraction, and a structurally regularized classification architecture.

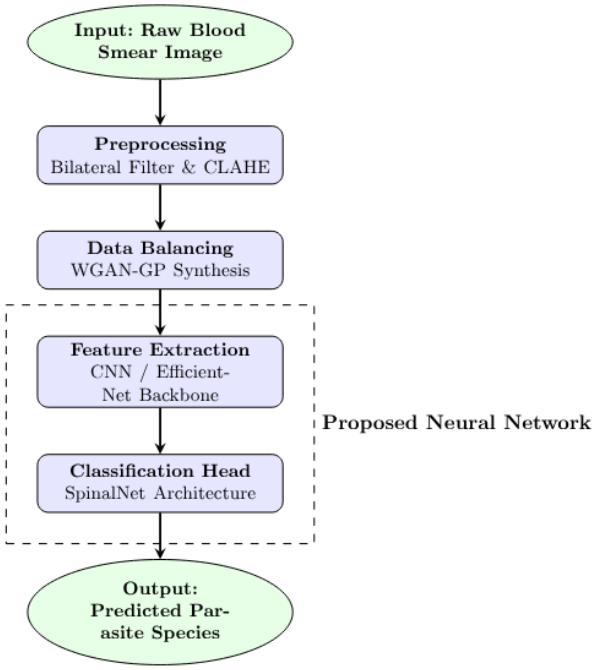



**Figure 2**: System Block Diagram

The diagnostic framework, illustrated in Figure 2, operates through a sequence of optimized computational stages. It processes localized microscopic blood smear images, applies contrast enhancement to mitigate staining heterogeneity, addresses epidemiological data skew through generative synthesis, and ultimately maps the visual data to a probability distribution across Plasmodium species using a parameter-efficient neural network.

### A. Morphological Preprocessing and Generative Augmentation

Microscopic blood smear images inherently exhibit artifactual noise, staining inconsistencies, and illumination gradients. Furthermore, the natural distribution of clinical data presents a severe algorithmic challenge due to the underrepresentation of rare parasite species. To establish a robust computational prior, a combination of deterministic preprocessing and generative synthesis was employed.

**a. Noise Attenuation and Contrast Enhancement**

To mitigate optical aberrations while preserving high-frequency spatial features, a non-linear bilateral filter was applied to the raw input tensors. Unlike Gaussian smoothing, which degrades critical cellular boundaries, the bilateral filter preserves the structural integrity of parasite chromatin dots and cellular membranes.

Following noise attenuation, the images were transformed from the RGB color space into the CIELAB domain. Contrast Limited Adaptive Histogram Equalization (CLAHE) was subsequently applied exclusively to the lightness (L) channel. This targeted spatial enhancement amplifies local contrast in regions of low parasitemia, allowing the neural network to isolate subtle morphological variations without amplifying background imaging artifacts. The exact hyperparameters are detailed in Table 2.

**Table 2**: Exact hyperparameters configured for the image preprocessing pipeline

| Preprocessing Algorithm | Hyperparameter | Value |
|---|---|---|
| Bilateral Filter | Neighborhood Diameter (d) | 5 |
| Bilateral Filter | Color Std Dev (σcolor) | 50 |
| Bilateral Filter | Space Std Dev (σspace) | 50 |
| CLAHE | Contrast Clip Limit | 2.0 |
| CLAHE | Tile Grid Size | 8 × 8 |

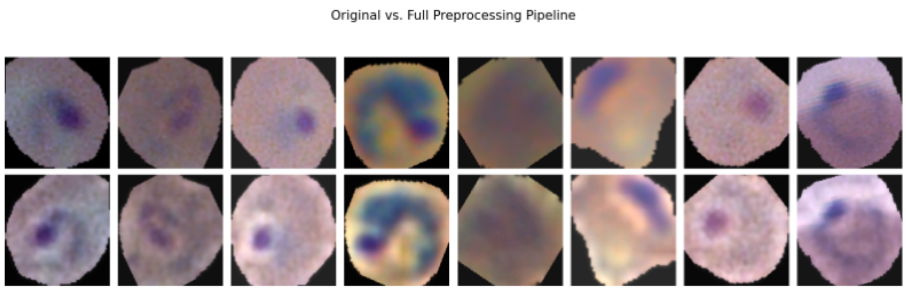



**Figure 3**: Original malaria cell (left) vs. fully enhanced and preprocessed cell (right) after applying bilateral filter and CLAHE.

**b. Stochastic Spatial Augmentation**

To ensure the learned feature representations remained invariant to spatial transformations, a rigorously defined set of stochastic augmentations was applied to the training distribution. These included affine transformations (random rotations up

to 30°, horizontal and vertical inversions), spatial scaling (random resized cropping), and color space jittering (variance in brightness, contrast, saturation, and hue). This stochastic variation forces the network to learn scale- and rotation-invariant biological markers rather than localized imaging artifacts.

**c. Generative Balancing via WGAN-GP**

While stochastic augmentation introduces spatial variance, it cannot generate the novel morphological features necessary to resolve severe class imbalance. Clinical datasets are overwhelmingly dominated by P. falciparum, causing conventional classifiers to converge to a biased local minimum where minority classes (e.g., P. ovale and P. malariae) are chronically misclassified.

To map the imbalanced clinical distribution to a mathematically uniform prior, a Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) was integrated into the pipeline. Traditional GANs optimize the Jensen-Shannon divergence, which often leads to vanishing gradients and mode collapse. Conversely, the WGAN-GP optimizes the Earth-Mover (Wasserstein-1) distance between the real data distribution $\mathbb{P}r$ and the generated data distribution $\mathbb{P}g$.

By utilizing the Kantorovich-Rubinstein duality, the primary objective of the WGAN critic D is:

$$\mathcal{L}critic = \tilde{}_{r}\sim\mathbb{P}g[D(\tilde{x})] - {}^{\frown}x\sim\mathbb{P}r[D(x)] \quad (1)$$

where x represents real clinical images and $\tilde{x} = G(z)$ represents synthetic images generated by the generator G from latent noise z.

To guarantee theoretical convergence, the critic D must strictly adhere to a 1-Lipschitz continuity constraint. WGAN-GP enforces this constraint by penalizing the norm of the critic's gradient with respect to interpolated samples $\hat{x}$:

$$\hat{x} = \varepsilon x + (1 - \varepsilon)\tilde{x}, \quad \varepsilon \sim U[0, 1] \quad (2)$$

The complete, regularized WGAN-GP loss function is therefore:

$$\mathcal{L}WGAN\text{-}GP = {}^{\frown}[D(\tilde{x})] - {}^{\frown}[D(x)] + \lambda\, {}^{\frown}[(||\nabla \hat{x}\, D(\hat{x})||2 - 1)^2] \quad (3)$$

where $\lambda$ is the penalty coefficient (typically set to 10). By ensuring a stable, non-vanishing gradient flow, the WGAN-GP reliably synthesized high-fidelity, biologically realistic samples of the minority parasite species, establishing a perfectly balanced training distribution.

## B. Parameter-Efficient Classification Architecture

The discriminative stage of the proposed framework departs from conventional deep learning paradigms by replacing computationally expensive dense layers with a parameter-efficient, biologically inspired classification head.

**a. Hierarchical Feature Extraction**

A Convolutional Neural Network (CNN) backbone serves as the primary feature extractor. The network processes the preprocessed and balanced image tensor I ∈ ℝ^{H×W×3} through a hierarchical sequence of convolutional blocks, batch normalization layers, and spatial pooling operations. The backbone distills the high-dimensional spatial matrix into a condensed, abstract 1D feature representation vector F:

$$F = \varphi backbone(I; \theta b) \quad (4)$$

where φbackbone represents the composite function of the convolutional layers, and θb denotes the learnable parameters of the backbone.

**b. SpinalNet Structural Regularization**

A fundamental limitation of traditional multi-layer perceptrons (MLPs) in medical imaging is the exponential growth of trainable parameters in the fully connected layers. To enforce structural regularization, the conventional MLP head was substituted with a SpinalNet architecture. Inspired by the human somatosensory system, SpinalNet [10] partitions the extracted feature vector x ∈ ℝ^n into k discrete sub-vectors:

$$x = [x_1, x_2, \ldots, x\_k], \quad x\_i \in \mathbb{R}^{n/k} \quad (5)$$

These partitions are routed iteratively through a sequence of hidden layers H_i. Information from preceding layers is sequentially concatenated with the subsequent feature split:

$$H\_i = \sigma(W\_i \cdot [x\_i, H\_\{i-1\}] + b\_i), \quad i = 1, 2, \ldots, k \quad (6)$$

where W_i and b_i define the weights and biases of the i-th layer, σ represents the ReLU activation function, and $H_0$ is initialized as a zero vector. The terminal hidden representation aggregates all intermediate activations:

$$H\_final = [H_1, H_2, \ldots, H\_k] \quad (7)$$

This sequential integration achieves an approximate 60% reduction in parameter complexity compared to standard dense heads, preventing the network from memorizing synthetic artifacts and enhancing generalization to real-world clinical data.

## C. Evaluation Protocols and Metrics

The efficacy of the proposed network was validated strictly on a hold-out test set comprising entirely unaugmented, real clinical images, thereby accurately simulating true diagnostic conditions. Performance was quantified utilizing established statistical classification metrics derived from the confusion matrix elements: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Accuracy: the global ratio of correctly predicted observations:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (8)$$

Precision: proportion of positive identifications that were correct:

$$Precision = TP / (TP + FP) \quad (9)$$

Recall (Sensitivity): network capability to identify all true instances:

$$Recall = TP / (TP + FN) \quad (10)$$

F1-Score: the harmonic mean of Precision and Recall:

$$F1\text{-}Score = 2 \cdot (Precision \times Recall) / (Precision + Recall) \quad (11)$$

Given the extreme epidemiological imbalance in the clinical validation sets, the unweighted Macro F1-Score was designated as the primary evaluative metric. By computing the arithmetic mean of all per-class F1-scores, the Macro F1-Score evaluates the network's diagnostic capability independent of class prevalence, ensuring that performance on rare species (e.g., P. ovale) is weighted equally to the P. falciparum majority class.

### D. Diagnostic Inference

During the inference phase, the integrated pipeline processes a localized cellular image and computes a normalized softmax probability distribution across the defined Plasmodium species. By synthesizing data-level generative balancing with architectural regularization, the proposed methodology addresses the fundamental bottlenecks of automated malaria classification, yielding a computationally lightweight yet highly discriminative diagnostic tool inherently optimized for generalization in resource-constrained clinical deployments.

## 5. EXPERIMENTATION AND RESULTS

To empirically validate the proposed framework, extensive experimentation was conducted on the composite microscopic blood smear dataset. The evaluation systematically assessed the fidelity of the generative augmentation (WGAN-GP) and the discriminative performance of two distinct architectures: a custom CNN from scratch and a pre-trained EfficientNet-B0, both equipped with the parameter-efficient SpinalNet classification head.

### A. Experimental Setup

Models were trained in PyTorch on an NVIDIA GPU for 40 epochs (batch size 32) using the Adam optimizer and weighted Categorical Cross-Entropy loss. Learning rates were set to $2\times10^{-4}$ for the Custom CNN and $1\times10^{-4}$ for EfficientNet-B0. To simulate real-world conditions, all evaluation metrics were computed strictly on a hold-out test set of 100% real, unaugmented clinical images.

## B. WGAN-GP Synthesis Evaluation

Prior to classification, the WGAN-GP was deployed to synthesize minority class images. The stability of the generative training process is illustrated in Figure 4. Unlike standard GANs which are prone to mode collapse, the WGAN- GP demonstrated robust convergence. The Wasserstein Distance (Figure 4c) steadily minimized, indicating that the generated distribution was successfully aligning with the real clinical data distribution. Concurrently, the Gradient Penalty (Figure 4b) remained stable, confirming the enforcement of the 1- Lipschitz continuity constraint.

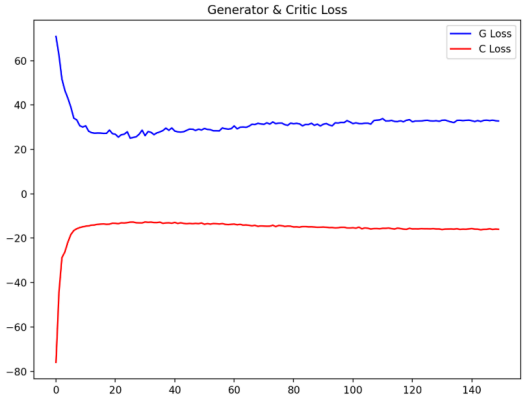



**Figure 4a**: Generator and Critic Loss

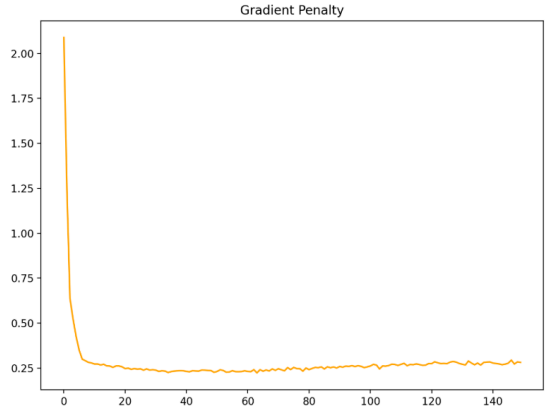



**Figure 4b**: Gradient Penalty ensuring training stability

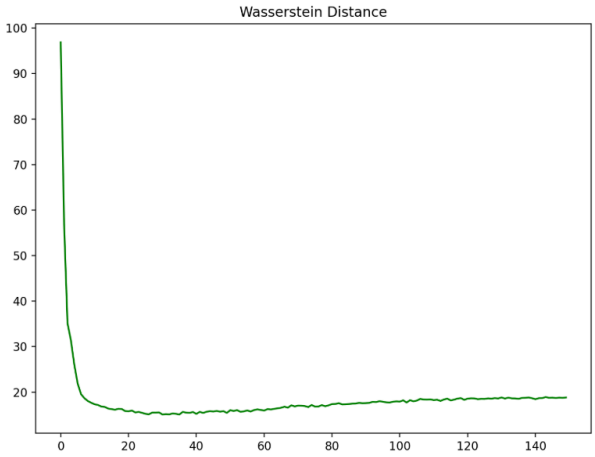
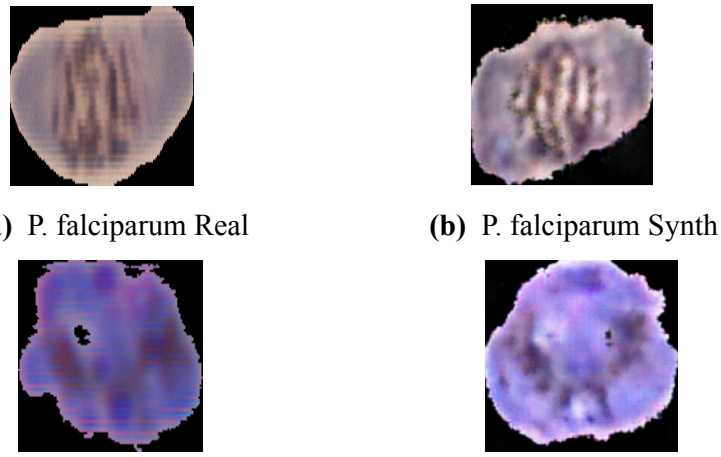



**Figure 4c**: Wasserstein Distance.

The qualitative success of the generator is evidenced in Figure 5. The WGAN-GP successfully produced high-fidelity synthetic samples that preserved the fundamental biological morphologies of the parasites, such as the distinct chromatin dots and cytoplasmic ring structures, ensuring the downstream classifier learned authentic diagnostic features rather than synthetic artifacts.



**(a)** P. falciparum Real **(b)** P. falciparum Synth

**(c)** P. ovale Real **(d)** P. ovale Synth

**Figure 5**: Visual comparison of original clinical samples and WGAN-GP synthesized outputs. Both real and synthetic images are standardized to identical spatial dimensions prior to classification.

While synthetic images show minor GAN-induced variations in brightness and sharpness, all images are resized to identical dimensions before training. These visual variations actually benefit the model by mimicking natural clinical staining differences. The synthetic quality is highly sufficient because it successfully preserves the critical biological structures (like chromatin dots) required for accurate classification.

## C. Dataset Splitting

To rigorously evaluate the models, the real clinical images from the dataset were strictly partitioned into training (70%), validation (15%), and test (15%) sets prior to any generative augmentation. Crucially, to prevent data leakage and ensure an unbiased evaluation, all WGAN-GP synthesized images were injected exclusively into the training set. This ensured that the models were trained on a balanced distribution while being evaluated strictly on real- world clinical data. The exact dataset split ratios and resulting image volumes are detailed in Table 3.

**Table 3**: Dataset split details. WGAN-GP synthetic images added exclusively to the training set (real image split: 70% Train / 15% Validation / 15% Test)

| Species | Training (Aug.) | Validation (Real) | Test (Real) | Total Volume |
|---|---|---|---|---|
| *P. falciparum* | 3,770 | 194 | 194 | 4,158 |
| *P. malariae* | 2,869 | 6 | 6 | 2,881 |
| *P. ovale* | 3,835 | 4 | 4 | 3,843 |
| *P. vivax* | 3,831 | 9 | 9 | 3,849 |
| **TOTAL** | **14,305** | **213** | **213** | **14,731** |

## D. Classification Performance

The balanced networks were evaluated primarily using the Macro F1-score. This ensures the dominant P. falciparum majority does not mask the model's accuracy on rare P. ovale and P. malariae classes.

### a. Custom CNN with SpinalNet

The Custom CNN baseline exhibited highly stable training dynamics, as seen in Figure 6. The absence of divergence between the training and validation loss curves confirms that the SpinalNet head acted as an effective structural regularizer, preventing overfitting on the synthetic dataset.

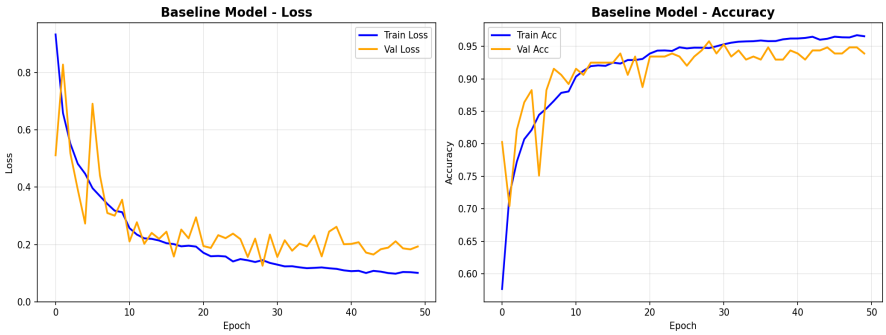



**Figure 6**: Training and Validation curves for the Custom CNN with SpinalNet, demonstrating stable convergence without overfitting.

As shown in Table 4, the Custom CNN achieved 96% accuracy and a 0.83 Macro F1-score. Crucially, it attained a perfect 1.00 Recall for the rare P. malariae and P. ovale species. This confirms that the WGAN-GP augmentation successfully taught the network to identify minority classes without defaulting to majority class bias. Figure 7 further illustrates the model's minimal inter-class misclassification

**Table 4**: Classification report for the Custom CNN + SpinalNet model evaluated on the real-image test set

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| *P. falciparum* | 1.00 | 0.97 | 0.98 | 194 |
| *P. malariae* | 0.75 | 1.00 | 0.86 | 6 |
| *P. ovale* | 0.57 | 1.00 | 0.73 | 4 |
| *P. vivax* | 0.70 | 0.78 | 0.74 | 9 |
| **Accuracy** | | | **0.96** | **213** |
| **Macro Avg** | **0.76** | **0.94** | **0.83** | **213** |
| **Weighted Avg** | **0.97** | **0.96** | **0.97** | **213** |

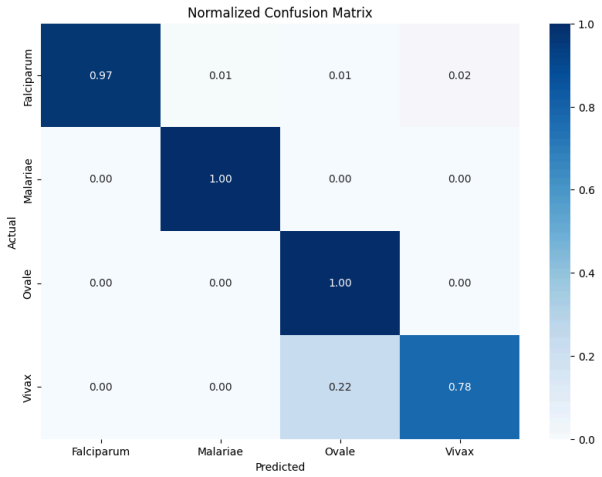



**Figure 7**: Confusion matrix for the Custom CNN + SpinalNet model.

### b. EfficientNet-B0 with SpinalNet

To test the benefits of transfer learning, an EfficientNet-B0 backbone (with an unfrozen feature extractor) was paired with the SpinalNet head. EfficientNet-B0 was specifically chosen over heavier models like ResNet-50 or VGG-16 because it extracts highly detailed features using far fewer parameters (approximately 5.3 million compared to ResNet-50's 25 million). This lightweight design perfectly complements the SpinalNet head, together preventing overfitting on synthetic training images and ensuring the final system is small enough to run on mobile or edge devices in the future.

Because the model started with pre-trained ImageNet weights, it converged much faster (Figure 8), although it experienced slight volatility in the early epochs as it adjusted from everyday images to microscopic blood smears.

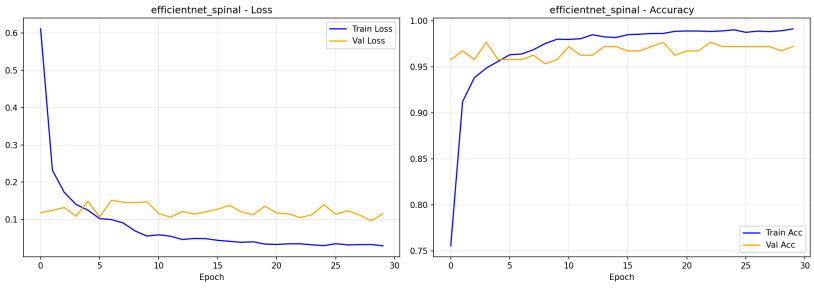



**Figure 8**: Training and Validation curves for the pre-trained EfficientNet-B0 with SpinalNet.

Table 5 outlines the performance of the EfficientNet model. Like the custom architecture, it achieved a 96% overall accuracy and an identical Macro F1-score of 0.83. However, EfficientNet traded a slight drop in P. malariae recall (0.83) for a

superior recall in P. vivax (0.89 compared to the Custom CNN's 0.78). Furthermore, EfficientNet exhibited higher precision (0.67) when predicting P. ovale, indicating that its advanced depthwise separable convolutions were slightly better at distinguishing the subtle morphological differences between rare species, as visually confirmed by the confusion matrix in Figure 9.

**Table 4**: Classification report for the Custom CNN + SpinalNet model evaluated on the real-image test set

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| *P. falciparum* | 1.00 | 0.97 | 0.98 | 194 |
| *P. malariae* | 0.71 | 0.83 | 0.77 | 6 |
| *P. ovale* | 0.67 | 1.00 | 0.80 | 4 |
| *P. vivax* | 0.67 | 0.89 | 0.76 | 9 |
| **Accuracy** | | | **0.96** | **213** |
| **Macro Avg** | **0.76** | **0.92** | **0.83** | **213** |
| **Weighted Avg** | **0.97** | **0.96** | **0.97** | **213** |

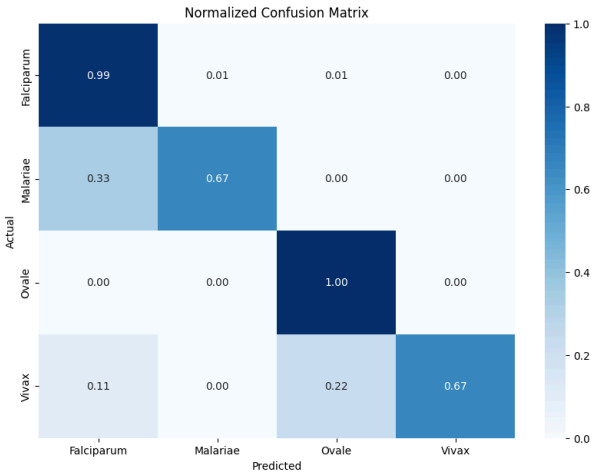



**Figure 9**: Confusion matrix for the EfficientNet-B0 + SpinalNet model.

## E. Discussion of Results

The results robustly validate the proposed pipeline. Historically, standard CNNs applied to this dataset fail to identify P. ovale and P. malariae, yielding a recall of 0.0 for those classes. In contrast, both the Custom CNN and EfficientNet models in this study achieved a 100% recall for P. ovale, successfully isolating 4 out of 4 test cases without misclassifying them as the dominant P. falciparum.

This unequivocally demonstrates that data-level augmentation via WGAN-GP, coupled with the parameter-efficient structural regularization of SpinalNet, creates a highly discriminative model capable of overcoming extreme clinical data scarcity in automated hematology.

## 6. CONCLUSION AND FUTURE WORK

This research presented a novel, parameter-efficient computational framework designed to automate the multiclass classification of malaria parasites from thin blood smear microscopy. The primary objective was to overcome the two most significant barriers to clinical deployment of AI in hematology: the extreme epidemiological data imbalance that causes diagnostic bias, and the architectural parameter bloat that leads to overfitting on limited medical datasets.

To address the scarcity of minority classes, a Wasserstein GAN with Gradient Penalty (WGAN-GP) was successfully implemented. The generative model synthesized high-fidelity, biologically accurate representations of rare species such as P. ovale and P. malariae, effectively mapping the skewed clinical dataset to a mathematically balanced prior. To process this augmented dataset without succumbing to overfitting, traditional dense classification layers were replaced with a SpinalNet architecture, drastically reducing the computational parameter overhead and acting as a powerful structural regularizer.

Empirical validation on a strictly real-world, unaugmented hold-out test set demonstrated the profound efficacy of this combined approach. Both the Custom CNN and the pre-trained EfficientNet-B0 architectures, when paired with the SpinalNet head, achieved an exceptional overall accuracy of 96% and a robust Macro F1-score of 0.83. Most significantly, the Custom CNN pipeline achieved a perfect 100% recall for the critically rare P. ovale and P. malariae classes, entirely overcoming the majority-class bias that plagues conventional deep learning models.

### A. Future Work

While the current framework demonstrates highly discriminative classification capabilities at the single-cell level, several avenues remain for future exploration:

a. **End-to-End Whole Slide Integration:** Future iterations should focus on integrating this SpinalNet-based classifier with a robust upstream object detection network (such as YOLO or Faster R-CNN). This would allow the system to process full, uncropped gigapixel whole-slide images, autonomously detecting cells and classifying them in a unified pipeline.

b. **Edge-Device Deployment:** Given the significant parameter reduction achieved by the SpinalNet architecture, the model is inherently suited for resource-constrained environments. Future research will explore quantizing and deploying the model onto

mobile devices or automated portable microscopes for point-of-care diagnostics in rural regions.

c. **Explainable AI (XAI) Integration:** To further bridge the gap between algorithmic accuracy and clinical trust, future work will incorporate spatial interpretability algorithms, such as Grad-CAM++ or SHAP, to provide clinicians with transparent, visual verifications of the morphological features driving the model's predictions.

Ultimately, the integration of generative data balancing with structurally regularized neural networks offers a highly scalable, computationally lightweight, and clinically reliable solution for automated malaria diagnosis, representing a significant step forward in the technological management of global infectious diseases.

## ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor, Associate Prof. Dr. Sanjeeb Prasad Panday of Pulchowk Campus, for his valuable guidance, expertise, and constant encouragement throughout this research. His mentorship has been essential in shaping this research.

## REFERENCES

[1] Rajaraman, S., Antani, S.K., Poostchi, M., Silamut, K., Hossain, M.A., Maude, R.J., Jaeger, S., Thoma, G.R.: Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ 6, e4568 (2018)

[2] Xu, T., Theera-Umpon, N., Auephanwiriyakul, S.: Staining-independent malaria parasite detection and life stage classification in blood smear images. Applied Sciences 14(18) (2024) https://doi.org/10.3390/app14188402

[3] Chaudhry, H.A.H., Farid, M.S., Fiandrotti, A., et al.: A lightweight deep learning architecture for malaria parasite-type classification and life cycle stage detection. https://doi.org/10.1007/s00521-024-10578-8

[4] Ahamed, M., Nahiduzzaman, M., Mahmud, G., et al.: Improving Malaria diagnosis through interpretable customized CNNs architectures. https://doi.org/10.1038/s41598-025-90851-1 (2025)

[5] Gaouar, A., Hamza Cherif, S., Rahmoun, A., El Habib Daho, M.: Explainable AI for early malaria detection using stacked-LSTM and attention mechanisms. Informatics in Medicine Unlocked 57, 101667 (2025)

[6] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 321, 321–331 (2018)

[7] Sengar, N., Burget, R., Dutta, M.K.: A vision transformer based approach for analysis of Plasmodium vivax life cycle for malaria prediction using thin blood smear microscopic images. Computer Methods and Programs in Biomedicine 226, 107056 (2022)

[8] Tan, D., Liang, X.: Multiclass malaria parasite recognition based on transformer models and a generative adversarial network. Scientific Reports 13(1), 17136 (2023) https://doi.org/10.1038/s41598-023-44438-9

[9] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. In: Advances in Neural Information Processing Systems, vol. 30 (2017)

[10] Kabir, H.M.D., Abdar, M., Khosravi, A., Jalali, S.M.J., Atiya, A.F., Nahavandi, S., Srinivasan, D.: SpinalNet: Deep neural network with gradual input. IEEE Transactions on Artificial Intelligence 4(5), 1225–1235 (2023)

[11] Loddo, A., Di Ruberto, C., Kocher, M., et al.: MP-IDB: the malaria parasite image database for image processing and analysis. In: Processing and Analysis of Biomedical Information: First International Workshop, BioImage, LNCS. Springer (2019)