PlasmoStage: A Hierarchical Deep Learning Framework for Plasmodium Parasite Staging in Malaria

1st Mostafa Salem
Department of Computer Vision
MBZUAI & Assiut University
Abu Dhabi, UAE & Assiut, Egypt
mostafa.salem@mbzuai.ac.ae

2nd Saad Abouzahir
Department of Computer Vision
MBZUAI
Abu Dhabi, UAE
abouzahir.saad@mbzuai.ac.ae

3rd Hosni Ghedira
Department of Computer Vision
MBZUAI
Abu Dhabi, UAE
Hosni.Ghedira.ac.ae

4rd Abdulmotaleb El Saddik
Department of Computer Vision
MBZUAI
Abu Dhabi, UAE
a.elsaddik@mbzuai.ac.ae

5th Mohammad Yaqub
Department of Computer Vision
MBZUAI
Abu Dhabi, UAE
mohammad.yaqub@mbzuai.ac.ae

***Abstract*—Malaria remains one of the most life-threatening infectious diseases worldwide. Recent advancements in deep learning and computer vision have shown significant promise in automating the analysis of medical images, including malaria-infected blood smears. While hierarchical classification, a machine learning approach that organizes classes into a hierarchy from broad categories to specific subtypes, has proven effective in various domains, its application to malaria parasite staging at the single-cell level remains underexplored. In this work, we present PlasmoStage, a novel hierarchical deep learning framework for hierarchical malaria parasite staging. Our approach leverages the DinoBloom foundation model, a state-of-the-art self-supervised model for single-cell image analysis in hematology, as a robust feature extractor. By fine-tuning these features using fully connected layers, PlasmoStage surpasses traditional flat classification models and baseline hierarchical methods. Our contributions are threefold: (1) We introduce a biologically inspired hierarchical classification framework that improves diagnostic accuracy and interpretability by aligning with the natural progression of malaria parasites. (2) We demonstrate the efficacy of foundation model-based feature extraction, achieving state-of-the-art performance with minimal fine-tuning. (3) We provide a comprehensive evaluation on publicly available datasets, including ablation studies and benchmarking against existing methods. Experimental results demonstrate that PlasmoStage effectively differentiates between uninfected and infected cells (Accuracy = 98.66%, F1-score = 98.38%), accurately identifies parasite species—Plasmodium falciparum or Plasmodium vivax (Accuracy = 98.75%, F1-score = 98.99%), and outperforms conventional flat and hierarchical classification approaches in parasite staging (Vivax staging: Accuracy = 85.90%, F1-score = 85.71%; Falciparum staging: Accuracy = 96.92%, F1-score = 96.62%). This work highlights the potential of leveraging foundation models for automated malaria diagnosis and staging.**

***Index Terms*—Malaria Parasite, Hierarchical Classification, Foundation Model, Deep Learning**

I. Introduction

Malaria remains one of the most life-threatening infectious diseases worldwide. According to the latest World Malaria report, there were 263 million cases of malaria in 2023 compared to 252 million cases in 2022. The estimated number of malaria deaths was 597,000 in 2023 compared to 600,000 in 2022 [1]. Malaria is a life-threatening disease that is spread to humans by some types of mosquitoes. There are four species of Plasmodium that are known to cause malaria in humans, with P. falciparum and P. vivax posing the most significant health threats. P. falciparum is the most lethal malaria parasite and is highly prevalent on the African continent, whereas P. vivax is the dominant species in most regions outside of sub-Saharan Africa. The remaining three species capable of infecting humans are P. malariae and P. ovale, the latter of which is a zoonotic parasite that can be transmitted from macaques to humans [2].

The accurate and timely diagnosis of malaria is critical for effective treatment and disease management. Traditional diagnostic methods, such as microscopy and rapid diagnostic tests (RDT), are widely used but suffer from limitations, including reliance on expert microscopists, variability in interpretation, and limited sensitivity for low parasitemia cases [3]. Recent advances in deep learning and computer vision have shown great promise in automating the analysis of medical images, including malaria blood smears. Convolutional neural networks (CNN) and transformer-based models have been successfully applied to tasks such as parasite detection, species classification, and identification of stages of life [4], [5]. However, most existing approaches treat malaria classification as a flat, multi-class problem, ignoring the inherent hierarchical structure of the task. For instance, a blood cell image

must first be classified as infected or uninfected, followed by species identification (P. falciparum or P. vivax), and finally, the parasite's developmental stage (ring, trophozoite, schizont, or gametocyte). Hierarchical classification not only aligns with the biological reality of malaria, but also improves the interpretability and performance of the model by taking advantage of the relationships between classes [6].

In this work, we propose a hierarchical deep learning framework for the classification of malaria parasites from single blood cell images. Our approach leverages the DinoBloom foundation model, a state-of-the-art self-supervised foundation model for single cell images in hematology, as a feature extractor. The extracted features are fine-tuned using fully connected layers before being passed to a hierarchical classification pipeline. The pipeline consists of three stages: 1) binary classification of cells as infected or uninfected, 2) species classification for infected cells (P. falciparum or P. vivax), and 3) stage classification for each species (ring, trophozoite, schizont, or gametocyte).

By incorporating hierarchical relationships into the classification process, our method achieves superior performance compared to flat classification approaches. The contributions of this work are threefold.

1) We introduce a hierarchical classification framework tailored to the biological structure of malaria parasites, allowing a more accurate and interpretable diagnosis.
2) We demonstrate the effectiveness of a foundation model as a feature extractor for malaria image analysis, achieving state-of-the-art performance with minimal fine-tuning.
3) We provide a comprehensive evaluation of our approach in publicly available datasets, including ablation studies and comparisons with existing methods.

The rest of this paper is structured as follows: Section II provides a review of related work on malaria classification and hierarchical deep learning. Section III details the datasets, model architecture, and training methodology. Section IV presents the experimental results and discussion. Finally, Section V summarizes the conclusions of the study.

II. Related Work

A. Malaria Parasite Classification

The automation of malaria diagnosis using machine learning and deep learning techniques has been an active area of research over the past decade. Early approaches relied on hand-crafted features extracted from blood smear images, such as texture, color, and shape descriptors, combined with traditional machine learning classifiers such as support vector machines (SVMs) and random forests [7], [8]. Although these methods demonstrated promising results, their performance was limited by the quality of the hand-crafted features and their inability to generalize across diverse datasets.

With the advent of deep learning, CNNs have become the defacto standard for malaria image analysis. Rajaraman et al. [9] proposed a CNN-based model to classify malaria-infected cells, achieving high precision in the publicly available Malaria Cell Image Dataset. Similarly, Liang et al. [4] developed a deep learning framework to detect and segment malaria parasites in blood smear images, demonstrating the potential of CNNs for automated diagnosis. More recently, transformer-based models, such as Vision Transformers (ViTs), have been applied to medical image analysis tasks, including malaria classification, due to their ability to capture long-range dependencies and global context [10].

Despite these advancements, most existing methods treat malaria classification as a flat multi-class problem, ignoring the inherent hierarchical structure of the task. For instance, a blood cell image must first be classified as infected or uninfected, followed by species identification (P. falciparum or P. vivax), and finally, the parasite's developmental stage (ring, trophozoite, schizont, or gametocyte). This flat classification approach not only fails to take advantage of the relationships between classes but also limits the interpretability and clinical utility of the results.

B. Hierarchical Deep Learning

Hierarchical classification is a well-studied problem in machine learning, with applications in domains such as text classification, image recognition, bioinformatics and leukemia [6], [11], [12]. The key idea is to organize classes into a hierarchy, where higher-level classes represent broader categories, and lower-level classes represent more specific subcategories. By incorporating hierarchical relationships into the classification process, models can improve their performance and interpretability. In the context of deep learning, hierarchical classification has been successfully applied to various tasks, including medical image analysis. For example, Li et al. [13] proposed a hierarchical CNN for skin lesion classification, where the model first predicts the general type of lesion (e.g., benign or malignant) and then refines the prediction to a specific subtype (e.g., melanoma or nevus). However, the application of hierarchical deep learning to the classification of malaria parasites remains underexplored. Most existing methods focus on flat classification, treating each class as independent and ignoring the relationships between them. This limits their ability to capture the biological structure of malaria parasites and hinders their performance in real-world scenarios.

C. Foundation Models for Feature Extraction

Foundation models have emerged as powerful tools for feature extraction in various domains, including natural language processing and computer vision [14]. These models are pre-trained on large-scale datasets using self-supervised learning techniques, enabling them to learn rich, generalizable representations that can be fine-tuned for specific tasks. In the context of medical image analysis, foundation models have shown great promise for tasks such as disease diagnosis, organ segmentation, and image retrieval [15]. The use of foundation models for malaria parasite classification is still in its early stages. Most existing methods rely on CNNs or ViTs trained

from scratch on small, domain-specific datasets, which limits their generalization ability. By leveraging a foundation model like DinoBloom [16] as a feature extractor, our approach addresses this limitation and achieves state-of-the-art performance with minimal fine-tuning.

III. Methodology

A. Dataset

Our study utilizes two publicly available datasets: 1) Broad Bioimage Benchmark Collection (BBBC041v1): This dataset contains thin blood smear images of Plasmodium vivax-infected human blood samples, including both healthy cells and cells infected with P. vivax parasites [2]. 2) MP-IDB (Malaria Parasite Image Database): This dataset provides thin blood smear images with healthy cells as well as cells infected by Plasmodium vivax and Plasmodium falciparum parasites, offering a more diverse set of samples for training and analysis [3]. These datasets form the foundation for training and validating our hierarchical classification model. The images are annotated by expert parasitologists and categorized into the following classes:

1) Uninfected: Blood cells without any signs of malaria infection.
2) Infected: Blood cells infected with malaria parasites, further divided into:
   - Plasmodium falciparum: Subdivided into ring, trophozoite, schizont, and gametocyte stages.
   - Plasmodium vivax: Subdivided into ring, trophozoite, schizont, and gametocyte stages.

The dataset is highly imbalanced, with a predominance of healthy cells and an overrepresentation of the ring stage in both Plasmodium vivax and Plasmodium falciparum, as well as the trophozoite stage in P. vivax. Fig. 1 illustrates representative samples of extracted cells, showcasing healthy cells, P. vivax stages (Ring, Trophozoite, Schizont, Gametocyte), and P. falciparum stages (Ring, Trophozoite, Schizont, Gametocyte).

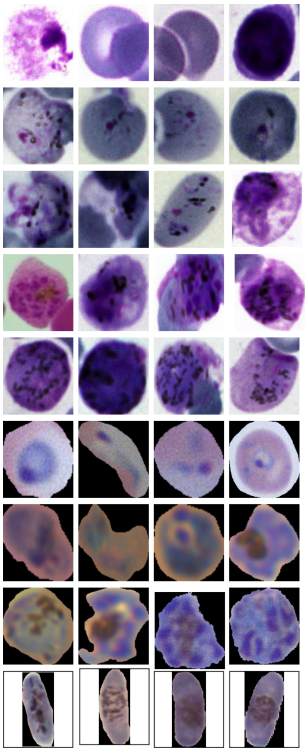

Fig. 1. Examples of extracted cells: The first row shows healthy cells. Rows 2-5 display the stages of Plasmodium vivax (Ring, Trophozoite, Schizont, and Gametocyte, respectively). Rows 6-9 illustrate the stages of Plasmodium falciparum (Ring, Trophozoite, Schizont, and Gametocyte, respectively).

B. Hierarchical Classification Pipeline

The overall model architecture is illustrated in Fig. 2. The proposed model leverages DinoBloom, a foundation model pre-trained on single-cell hematological images from diseases unrelated to malaria, as a feature extractor. In this transfer learning setup, the pre-trained weights of DinoBloom are initially frozen to preserve the generalizable representations learned from diverse hematological data. The model is then fine-tuned on the malaria dataset, allowing it to adapt these features to the specific task of malaria parasite classification. This approach effectively combines the broad generalization capability of DinoBloom with task-specific adaptation, enhancing classification performance while mitigating overfitting. The extracted features are subsequently passed through a series of fully connected layers that are fine-tuned to optimize the hierarchical classification pipeline.

This pipeline is structured into three stages, reflecting the inherent biological hierarchy of malaria infection. First, the model classifies the cell as either uninfected or infected. If the cell is infected, the model then determines the parasite species (P. falciparum or P. vivax). Finally, the developmental stage of the parasite (ring, trophozoite, schizont, or gametocyte) is identified. This hierarchical classification approach not only aligns with the biological progression of malaria infection but also enhances model interpretability and performance by leveraging the dependencies between classes and reducing error propagation.

- **Stage 1 (Infection Head – Uninfected vs. Infected Classification):** This stage consists of a binary classification layer with a sigmoid activation function, which determines whether the input cell is uninfected or infected. The output probability guides the subsequent classification stages, ensuring that only infected cells proceed for further analysis.
- **Stage 2 (Parasite Species Head – Plasmodium falciparum vs. Plasmodium vivax Classification):** If the cell is classified as infected in Stage 1, the extracted feature vector is forwarded to a second binary classification layer with a sigmoid activation function. This layer predicts the parasite species, distinguishing between P. falciparum and P. vivax.
- **Stage 3, 4 (Parasite Developmental Stage Classification – Ring, Trophozoite, Schizont, Gametocyte):** Based on the species identified in Stage 2, the feature vec-

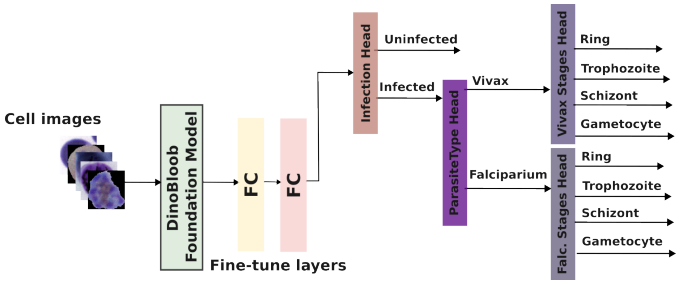

Fig. 2. Architecture of the proposed model for hierarchical malaria parasite classification, featuring multiple classification heads for infection status (uninfected/infected), parasite type (Plasmodium vivax/Plasmodium falciparum), and parasite stage (Ring, Trophozoite, Schizont, Gametocyte).

tor is processed through a multi-class classification layer with a softmax activation function. This layer predicts the parasite's developmental stage, classifying it as ring, trophozoite, schizont, or gametocyte. The hierarchical structure of the classification ensures biologically consistent predictions while leveraging the relationships between infection status, species, and developmental stages.

C. Implementation Details

The proposed model was trained and evaluated using stratified K-fold cross-validation with $K = 5$ to maintain balanced class distributions across all folds. The dataset was partitioned into five folds, where each fold served as the test set once, while the remaining four folds were used for training and validation. Within each training fold, data was split into training (80%) and validation (20%) subsets, ensuring robust and reliable results by mitigating potential biases from data splits. To address the class imbalance present in the dataset, we employed multiple strategies. First, stratified sampling was used during the training-validation splits to preserve representative distributions of each class. Second, targeted data augmentation techniques—such as rotation, flipping, and color jittering—were applied to increase the diversity and quantity of samples for minority classes.

Feature Extraction and Model Architecture. We utilized the DinoBloom foundation model as a feature extractor, leveraging its pre-trained weights, which were kept frozen during training to preserve the learned generalizable features. To adapt these features for malaria parasite classification, we added two fully connected layers with 512 and 128 units, respectively, each followed by ReLU activation functions. The classification layers were organized according to our three-stage hierarchical structure, fine-tuning these layers to optimize the hierarchical classification pipeline.

Optimization and Loss Functions. The model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. Given the hierarchical nature of the task, we employed different loss functions at each classification stage:

- **Stage 1 (Infection Classification):** Binary cross-entropy loss for classifying cells as uninfected or infected.
- **Stage 2 (Species Classification):** Binary cross-entropy loss to distinguish between *Plasmodium falciparum* and *Plasmodium vivax*.
- **Stage 3 (Developmental Stage Classification):** Categorical cross-entropy loss for classifying the specific developmental stage (ring, gametocyte, schizont, or trophozoite) of each species.

Training Strategy. The training process was monitored using validation accuracy. Early stopping was applied to prevent overfitting and improve generalization. All experiments were implemented using the PyTorch framework and conducted on an NVIDIA A100 GPU, ensuring efficient training and evaluation.

D. Ablation Study

To assess the impact of various components in our hierarchical classification framework, we conducted an ablation study by systematically modifying key elements of our model as follows:

- **Zero-Shot Classification with DinoBloom's Learned Representations:** To evaluate the effectiveness of fine-tuning DinoBloom's extracted features, we tested a zero-shot classification approach using its pre-trained feature representations without additional fine-tuning:
  - Embeddings were extracted from DinoBloom, followed by a prototype-based zero-shot classification approach.
  - Class prototypes were generated for each classification head (Infection Head, Parasite Species Head, and Parasite Stage Head) by computing the mean embedding of labeled reference samples per class.
  - Classification was performed using cosine similarity, assigning input embeddings to the class with the highest similarity to its corresponding prototype.
  - Infection classification involved comparing cell embeddings to uninfected and infected prototypes.
  - For infected cells, cosine similarity determined the parasite species (*Plasmodium falciparum* or *Plasmodium vivax*).
  - Finally, embeddings of correctly classified *P. falciparum* and *P. vivax* cells were matched to their respective stage prototypes (four stages per species).

  This hierarchical zero-shot method enabled classification without explicit fine-tuning of the model.
- **Training Baseline Models from Scratch:** We trained ResNet50 [17] and ViT-base [10] from scratch to evaluate their capacity to learn from limited malaria datasets without the benefit of prior knowledge from large-scale image corpora.
- **Flat vs. Hierarchical Classification:** To assess the importance of the hierarchical classification strategy, we replaced it with a flat classification approach where all classes (uninfected, parasite type, *P. falciparum* stages, and *P. vivax* stages) were predicted in a single step. This experiment provided insights into the effectiveness of the

TABLE I

PERFORMANCE COMPARISON OF THE PROPOSED MODEL ACROSS HIERARCHICAL AND FLAT CLASSIFICATION APPROACHES. THE TABLE REPORTS ACCURACY, AUC-ROC, RECALL, PRECISION, AND F1-SCORE, WITH AUC-ROC, RECALL, PRECISION, AND F1-SCORE FOR MULTI-CLASS HEADS CALCULATED AS WEIGHTED AVERAGES BASED ON SUPPORT (THE NUMBER OF TRUE INSTANCES FOR EACH LABEL).

Head (Hierarchical or Flat)	Accuracy (%) ↑	AUC-ROC (%) ↑	Recall (%) ↑	Precision (%) ↑	F1-score (%) ↑
Infection Head					
DinoBloom (Zero-Shot)	93.85	98.25	95.18	89.81	92.42
Hierarchical					
DinoBloom (Fine-tuning)	98.66	99.67	98.81	98.34	98.38
ResNet50 [17]	81.00	79.99	75.21	77.39	75.79
ResNet50-WeightedLoss	97.44	99.16	96.64	96.83	96.73
ViT-base [10]	73.14	70.58	58.57	68.76	62.11
ViT-base-WeightedLoss	96.55	99.54	95.10	98.25	95.50
ViT-base-SeparateTokens	98.12	99.65	97.00	97.23	97.20
Flat					
DinoBloom (Fine-tuning)	98.54	99.63	97.53	98.24	98.13
ResNet50 [17]	90.12	94.67	89.14	86.90	87.77
ViT-base [10]	98.02	99.43	96.95	97.10	97.00
Parasite Type Head					
DinoBloom (Zero-Shot)	98.41	97.99	97.97	98.95	98.53
Hierarchical					
DinoBloom (Fine-tuning)	98.75	98.09	98.05	99.95	98.99
ResNet50 [17]	72.29	77.61	56.09	99.09	71.32
ResNet50-WeightedLoss	96.59	94.62	94.52	99.20	97.15
ViT-base [10]	57.32	64.48	35.52	90.77	46.94
ViT-base-WeightedLoss	97.16	96.50	95.44	99.00	97.67
ViT-base-SeparateTokens	98.20	98.03	97.75	98.89	98.40
Flat					
DinoBloom (Fine-tuning)	97.53	98.02	96.09	99.86	97.98
ResNet50 [17]	90.40	98.03	85.31	99.24	91.63
ViT-base [10]	97.29	98.00	95.67	98.80	97.78
Vivax Stages					
DinoBloom (Zero-Shot)	70.31	98.94	70.31	79.00	72.68
Hierarchical					
DinoBloom (Fine-tuning)	85.90	90.79	85.90	85.68	85.71
ResNet50 [17]	56.00	68.14	56.00	67.22	56.66
ResNet50-WeightedLoss	81.14	87.77	81.14	80.60	80.65
ViT-base [10]	38.13	54.00	38.13	53.01	38.22
ViT-base-WeightedLoss	78.22	87.17	78.22	78.30	78.25
ViT-base-SeparateTokens	82.88	90.18	82.88	83.08	82.90
Flat					
DinoBloom (Fine-tuning)	80.92	89.83	80.92	83.69	82.15
ResNet50 [17]	61.04	76.14	61.04	71.48	62.96
ViT-base [10]	79.39	89.41	79.39	82.47	80.45
Falciparum Stages					
DinoBloom (Zero-Shot)	89.51	92.73	89.51	96.42	92.27
Hierarchical					
DinoBloom (Fine-tuning)	96.92	91.77	96.92	96.59	96.62
ResNet50 [17]	93.09	73.85	93.09	91.09	91.85
ResNet50-WeightedLoss	95.61	82.57	95.61	93.94	94.22
ViT-base [10]	94.83	54.49	94.83	89.94	92.32
ViT-base-WeightedLoss	95.61	79.91	95.61	95.26	95.50
ViT-base-SeparateTokens	96.50	85.93	96.92	95.95	96.35
Flat					
DinoBloom (Fine-tuning)	95.91	88.23	95.91	95.58	95.66
ResNet50 [17]	76.10	63.92	76.10	90.33	82.38
ViT-base [10]	96.43	85.58	96.43	96.10	96.30

hierarchical strategy and underscored the advantages of leveraging feature extractors from pretrained foundation models for accurate malaria parasite classification.

IV. RESULTS AND DISCUSSION

Table I presents the performance comparison of the proposed model using the frozen pretrained feature extractors from DinoBloom model and ResNet50 and ViT-base models across hierarchical and flat classification approaches. The zero-shot classification results using DinoBloom model are also reported. The table reports Accuracy, AUC-ROC, Recall, Precision, and F1-score, where the latter four metrics for multi-class classification heads are computed as weighted averages based on class support (i.e., the number of true instances per class). The hierarchical DinoBloom-based model achieves the highest performance across all classification heads.

Fig. 3 displays the ROC curves for each classification head of the hierarchical DinoBloom-based model, illustrating the model's ability to distinguish between classes at different hierarchical levels. These curves provide a comprehensive visualization of the trade-off between sensitivity (true positive rate) and the false positive rate, offering valuable insights into the classification effectiveness of each head.

Our experiments reveal a significant contrast in classification performance when employing hierarchical versus flat classification across different feature extraction backbones. When using DinoBloom as a feature extractor, hierarchical classification consistently outperforms flat classification. This can be attributed to DinoBloom's self-supervised training, which inherently structures its feature space in a hierarchical manner by clustering semantically similar representations. As a result, hierarchical classification effectively aligns with this structured embedding space, allowing for stepwise decision-making (infection $\rightarrow$ parasite type $\rightarrow$ stage) with reduced class overlap. In contrast, flat classification with DinoBloom struggles due to the similarity of certain feature embeddings, where early-stage infections from different parasites may be closely clustered, leading to misclassifications when using a single-step classifier.

Conversely, when using ResNet or ViT, flat classification initially outperformed hierarchical classification, likely due to error propagation—misclassifications in the infection classification step led to incorrect downstream decisions in parasite and stage classification. To address this issue, we introduced weighted loss training, giving higher importance to the infection classification head to ensure more reliable initial predictions. By dynamically adjusting loss weights based on task difficulty, we successfully improved hierarchical classification performance for ResNet, reducing error propagation and making it more effective than flat classification in this case.

However, weighted loss training did not improve ViT's hierarchical classification. This discrepancy can be explained by fundamental differences in how ResNet and ViT extract features. ResNet, as a convolutional neural network, builds representations in a hierarchical manner—starting from low-level features (edges, textures) and progressively forming higher-level semantic concepts. This aligns well with hierarchical classification, as each classification step refines local details relevant to the specific decision level. By emphasizing the infection classification head, ResNet learns more discriminative features at earlier stages, improving overall performance.

In contrast, ViT lacks an inherent hierarchical feature extraction process. Instead of progressively building local-to-global representations, ViT processes the entire image globally from the start using self-attention mechanisms. As a result, ViT does not naturally benefit from the hierarchical classification structure. While weighted loss adjustments improved its performance, it still remained lower than that of the ViT flat classification. Errors in earlier decisions (e.g., infection classification) still propagate downstream, as ViT's global feature learning does not explicitly refine representations step by step. This explains why hierarchical classification, even with weighted loss, does not outperform flat classification in ViT.

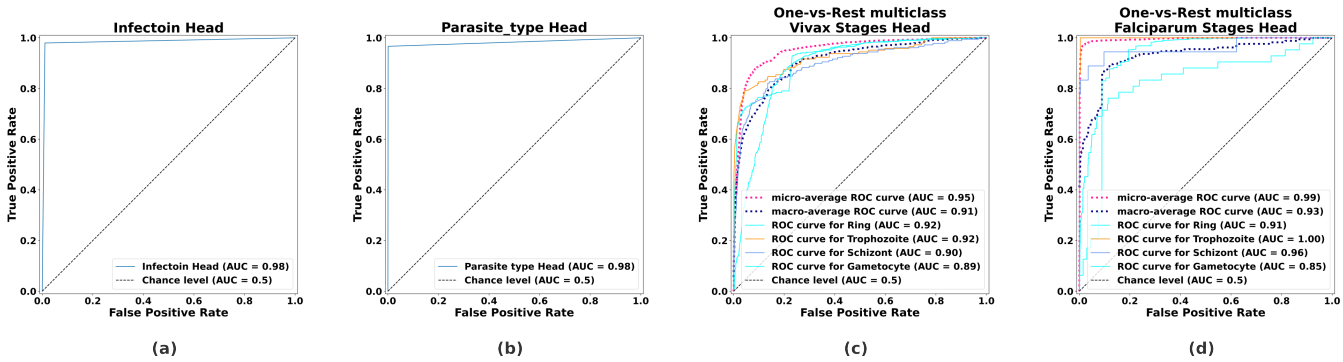

Fig. 3. ROC curves for the hierarchical classification heads: (a) infection classification (uninfected vs. infected), (b) parasite type classification (Plasmodium vivax vs. Plasmodium falciparum), (c) P. vivax stage classification, and (d) P. falciparum stage classification. Each curve represents the model's ability to distinguish between classes, highlighting its performance across different hierarchical levels.

To address this, we propose an alternative hierarchical token-based ViT approach, where instead of a single classification token, we introduce separate class tokens for infection, parasite type, and stage classification. This method allows ViT to independently learn distinct feature representations for each hierarchical level, mitigating error propagation issues. Our results demonstrate that this approach outperforms both the hierarchical ViT with a single class token and the flat ViT, highlighting the effectiveness of using separate class tokens for hierarchical classification.

Regarding the datasets, the two datasets may differ in imaging hardware and protocols, potentially introducing domain shifts. While we did not explicitly address this, using a foundation model helped mitigate such variability. We plan to explore domain adaptation strategies in future work.

V. Conclusion

In this study, we proposed a hierarchical deep learning framework for the classification of malaria parasites from single blood cell images. The framework utilizes the DinoBloom foundation model as a frozen feature extractor, leveraging its pre-trained representations to capture robust and generalizable features. These features are further refined using fully connected layers before being fed into a hierarchical classification pipeline. The model demonstrates high performance in distinguishing between uninfected and infected cells, identifying the species of the parasite (Plasmodium falciparum or Plasmodium vivax), and classifying the developmental stage (ring, trophozoite, schizont, or gametocyte). By reflecting the biological progression of malaria infection, the hierarchical classification approach not only improves accuracy but also enhances the interpretability of the model's predictions by leveraging the dependencies between classification stages.

References

[1] W. H. Organization, *World malaria report 2023*. World Health Organization, 2023.

[2] A. F. Cowman, J. Healer, D. Marapana, and K. Marsh, "Malaria: biology and disease," *Cell*, vol. 167, no. 3, pp. 610–624, 2016.

[3] A. Moody, "Rapid diagnostic tests for malaria parasites," *Clinical microbiology reviews*, vol. 15, no. 1, pp. 66–78, 2002.

[4] Z. Liang, A. Powell, I. Ersoy, M. Poostchi, K. Silamut, K. Palaniappan, P. Guo, M. A. Hossain, A. Sameer, R. J. Maude *et al.*, "CNN-based image analysis for malaria diagnosis," in *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2016, pp. 493–496.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need. advances in neural information processing systems," *Advances in neural information processing systems*, vol. 30, no. 2017, 2017.

[6] C. N. Silla and A. A. Freitas, "A survey of hierarchical classification across different application domains," *Data mining and knowledge discovery*, vol. 22, pp. 31–72, 2011.

[7] F. B. Tek, A. G. Dempster, and I. Kale, "Malaria parasite detection in peripheral blood images." BMVA, 2006.

[8] L. Rosado, J. M. C. Da Costa, D. Elias, and J. S. Cardoso, "Automated detection of malaria parasites on thick blood smears via mobile devices," *Procedia Computer Science*, vol. 90, pp. 138–144, 2016.

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

[10] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[11] I. Hamdi, H. El-Gendy, A. Sharshar, M. Saeed, M. Ridzuan, S. K. Hashmi, N. Syed, I. Mirza, S. Hussain, A. M. Abdalla *et al.*, "Breaking down the hierarchy: A new approach to Leukemia classification," in *International Workshop on Applications of Medical AI*. Springer, 2023, pp. 104–113.

[12] N. Syed, M. E. S. Saeed, S. Hussain, I. Mirza, A. M. Abdalla, E. A. Al Zaabi, I. Afrooz, S. Hashmi, and M. Yaqub, "Novel hierarchical deep learning models predict type of Leukemia from whole slide microscopic images of peripheral blood," *Journal of Medical Artificial Intelligence*, vol. 8, 2025.

[13] Y. Seo and K.-s. Shin, "Hierarchical convolutional neural networks for fashion image classification," *Expert systems with applications*, vol. 116, pp. 328–339, 2019.

[14] A. Thieme, A. Nori, M. Ghassemi, R. Bommasani, T. O. Andersen, and E. Luger, "Foundation models in healthcare: opportunities, risks & strategies forward," in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–4.

[15] R. J. Chen, T. Ding, M. Y. Lu, D. F. Williamson, G. Jaume, A. H. Song, B. Chen, A. Zhang, D. Shao, M. Shaban *et al.*, "Towards a general-purpose foundation model for computational pathology," *Nature Medicine*, vol. 30, no. 3, pp. 850–862, 2024.

[16] V. Koch, S. J. Wagner, S. Kazeminia, E. Sancar, M. Hehr, J. A. Schnabel, T. Peng, and C. Marr, "DinoBloom: a foundation model for generalizable cell embeddings in hematology," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 520–530.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.