AJPMHS GLOBAL RESEARCH NETWORK

AMERICAN Journal of Pediatric Medicine and Health Sciences
Volume 4, Issue 05, 2026 ISSN (E): 2993-2149

Hybrid CNN-VIT Model for Automated Multi-Stage Malaria Parasite Classification in Microscopy Images

Khalaf Hasan Taresh Hussein
Department of Biology, College of Education for Pure Sciences, Tikrit University, Iraq
E-mail: KHTpeps5@st.tu.edu.iq

Abstract: The diagnosis of malaria continues to be a big problem for treatment and surveillance of the disease in a timely manner, especially in places with limited resources where there are no expert microscopists. In this research, a new method that uses deep learning is presented combining EfficientNetB0-based convolutional neural network (CNN) with a Vision Transformer (ViT) in a very effective way to the automatic multi-stage malaria parasite classification from blood smear microscopy images. The local morphological features are extracted hierarchically by the CNN component, whereas the global spatial relationships and the long-range contextual dependencies within the infected cells are modeled by the ViT module enabling the improvement in the discrimination of the visually similar parasite stages. To increase the model's generalization between the different datasets and staining variabilities to even further extent, contrastive self-supervised learning (CSSL) has been integrated during the model training alongside morphology-aware data augmentation, so that the model can learn feature representations that are stage-consistent and stain-invariant. The results of the experiments reveal that the hybrid CNN–ViT model is a better performer than conventional deep learning architectures by a large margin, thus attaining 99.1% accuracy, 98.9% precision, 99.0% recall, and an F1-score of 98.9% on the validation set. These results underscore the integration of convolutional feature extraction and transformer-based global attention in the automation of malaria parasite analysis and point out that the proposed framework provides a trustworthy and scalable alternative to manual microscopy in clinical and field settings.

Keywords: Deep Learning, Vision Transformers, Malaria Parasite Classification, Microscopy Images, Convolutional Neural Networks.

Introduction

Parasitic infections continue to be a significant health problem all over the world, especially in underdeveloped areas where there is no access to specialized diagnostic methods [1]. Diseases by parasites such as malaria, babesiosis, and other blood-borne parasites need early and certain diagnosis in order to start treatment and avoid complications [2]. Parasitologists have always depended on manual examination of microscopy images of stained blood smears which, besides being a long process, are also subjected to inter-observer variability. The growing demand for fast, accurate, and fully automated diagnostic solutions has pushed the use of artificial intelligence (AI), especially deep learning, in medical image analysis [3], [4]. Deep learning, particularly convolutional neural networks (CNNs), has reached the best performance in all medical image applications, such as cell segmentation, classification, and disease diagnosis [5], [6]. However, most of the current deep learning methods for detecting parasitic infections are trained on small datasets and cannot be used for different imaging conditions and parasite species. Moreover, the unavailability of large annotated datasets and the difficulty in detecting parasite stages among blood components (for example, red blood cells, leukocytes) are prominent challenges that lead to low diagnostic accuracy [7], [8], [9].The main advancement of

this research is the creation of a hybrid convolution-transformer classification framework (Figure1) that brings together in a very explicit way the hierarchical morphological feature learning and global contextual reasoning for the multi-stage recognition of malaria parasites. The integration of a lightweight CNN backbone with a Vision Transformer module enables the proposed architecture to jointly capture the fine-grained cellular textures and long-range spatial dependencies, thus providing a more discriminative representation of the parasite stages than the conventional CNN-only models. In addition to this, the model's robustness under real-world acquisition variability is enhanced using contrastive self-supervised learning (CSSL) which is incorporated during training and encourages the model to learn stain-invariant and morphology-consistent feature embeddings. This design consequently allows the framework to effectively generalize across varied parasite stages, and staining conditions, and shifts in inter-dataset distribution, thus removing a major disadvantage of the existing supervised methods. There has been extensive experimental evaluation on several microscopy datasets which indicates that the proposed CNN-vision transformer (ViT) synergy consistently outperforms standard CNN architectures, especially at the difficult class boundaries where visually similar parasite stages are found. The results not only affirm the role of convolutional local feature extraction and transformer-based global attention as complementary to each other but also support their unified exploitation for automated malaria diagnosis as an effective strategy.

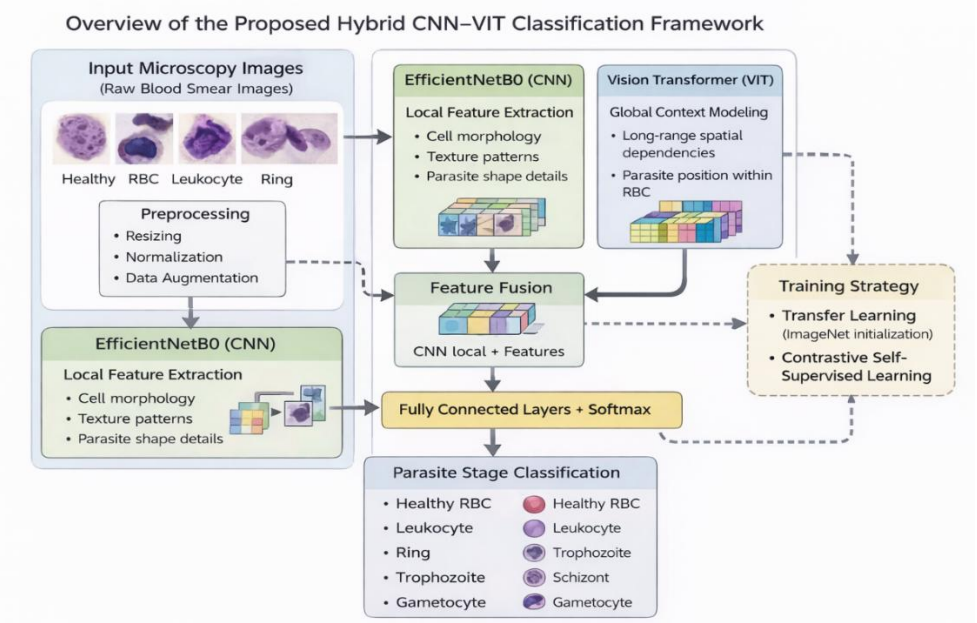

Figure 1. Flowchart of the proposed work.

Related work

Deep learning techniques for malaria diagnosis have made the biggest uplift in parasitic detection in microscope images, both accuracy-wise and efficiency-wise. Many researchers have investigated a variety of convolutional neural networks (CNNs) and object detection models for the progress of malaria detection. The CNN model for malaria detection was proposed in [10], which achieved an accuracy of 97.76%, using Kaggle's microscopic images, thus showing the promise of deep learning for parasitic detection automation. Similarly, another study [11] compared YOLOv5x, Faster R-CNN, SSD, and RetinaNet on a dataset of 2571 thick smear images, revealing that YOLOv5x achieved a recall of 93.50% and was used in a smartphone-based automated diagnostic system. The fully automated system incorporated AI-based parasite detection and robotic microscopy for auto-focusing and slide control, thus reducing human interaction to a minimum and efficiency to maximum. Besides detection, the morphological abnormalities of red blood cells (RBCs) have also been investigated as additional supportive

diagnostic features. A CNN-based classifier was proposed as a solution in [12] to classify the morphology of rouleaux (that is, stacked RBCs) and normal RBCs in thin blood smear images with an accuracy of 90.95%. This method not only detects the presence of malaria parasites but also reveals the associated RBC abnormalities through the use of microscopy-based diagnostics. Moreover, a dual-attention CNN model was presented in [13] for classification of leukocyte blood smears with the WBC identification and separation of Plasmodium-infected cells being enhanced. The DCGAN-based data augmentation was employed in the study to provide a better generalization of the model over PBC, LISC, and Raabin-WBC datasets with 99.83% accuracy, hence proving the attention-based deep learning models effectiveness in hematology applications. A number of scientists have concentrated their efforts on the classification of parasite species as a means to ease the updating of malaria treatment protocols. The PlasmodiumVF-Net that was suggested in [14] intermingled Mask R-CNN and ResNet50 and thus was able to get 90% plus accuracy in recognizing and classifying Plasmodium falciparum and Plasmodium vivax at the patient level in species-level infections. Similarly, [15] revealed a CNN model for classifying P. falciparum, P. vivax, and healthy cells which had an accuracy of 99.51%, and this shows the efficacy of deep learning for species-level diagnosis. An avant-garde YOLO-based model in [16] was also proposed for the multi-species parasite detection, which made its application to more than one parasite species and significantly increased the diagnostic performance [17], [18], [19]. Besides, the combination of SSL and transfer learning has been implemented to enhance malaria detection performance in low-resource settings [20], [21]. A malaria classification model based on VGG19 that went through exchange learning was put forward in [22], which was able to enable ecosystem-aware malaria detection and achieved over 90% accuracy at the same time. Furthermore, a YOLO-mp model was developed in [23] with the purpose of achieving the highest possible accuracy and speed trade-offs, which already outperformed YOLOv4 by a large margin with very high computational efficiency and the ability to be used on low-resource devices. There are still challenges to be faced in obtaining deep learning models that are robust, scalable, and generalizable for malaria detection under different imaging conditions. Several researches have pointed out the weakness of datasets, as the models trained on publicly available malaria datasets scarcely generalize to real clinical conditions [24]. In [25], a challenge was met with the invention of a mobile phone application-based malaria screening device that uses SSD multibox object detection and a VGG16-based classifier, which brought about rapid and affordable malaria detection in remote and resource-limited surroundings. Also, there have been studies focusing on hybrid deep learning models that integrate CNNs with attention mechanisms in order to raise the level of accuracy in parasite detection. The authors of [26] have come up with a CNN-based AI system that is capable of diagnosing malaria with 97.81% accuracy, reducing both the time and cost of diagnosis. A similar CNN-SVM hybrid model in [27] also managed to achieve 98% accuracy in the classification of four Plasmodium species, which is a strong indicator of the success of the combined use of deep learning and traditional machine learning methods for malaria diagnosis.The creative deep learning technique in [28] presented a deep transfer graph convolutional network (DTGCN) for the malaria parasite detection, which is the first application of the graph convolutional networks (GCN) to the multi-stage classification. By taking advantage of unsupervised learning for the movement of morphological information between the labeled source images, the DTGCN was able to achieve the performance metrics of 95.4% for accuracy, precision, recall, and F1-score.

Methods

1. Dataset

The data set [28],[29] from which this paper takes draws its information consists of a total of six classes of cells and of those two classes are uninfected (red blood cells and leukocytes) which are represented with color codes. The rest four classes are of the infected cells (gametocytes, rings, trophozoites, and schizonts) and each one has its color code. Microscopic images that have been

dyed with Giemsa stain are made more visible in the microscope view [30]. The original purpose of this dataset was not the identification of the precise stage of malaria parasite but rather the detection of the parasitized cells; still, the bounding box annotations and the stage labels allow for the implementation of both object detection and classification schemes. A total of 79,672 images (Figure 2) of both types of cells (parasitized and uninfected) were obtained from raw microscopy images by cropping based on the bounding box coordinates given. However, as mentioned in Supplementary Table S1, there is a very severe class imbalance, where the 5,000 selected RBC images alone account for 97.2% of the whole dataset. The class of leukocytes was the least represented one with just 103 images, and in order to increase the testing performance, 104 more self-acquired leukocyte images were added to the originally collected ones. To ensure an evaluation that was balanced across the board, the test dataset was made up of 600 images, which were generated by randomly picking 100 images from every class, and the rest of the images were used for training the model. The dataset for training comprised 6,856 images, among which were 44 gametocyte images, 107 leukocyte images, 5,000 RBC images, 253 ring images, 79 schizont images, and 1,373 trophozoite images. The dataset for testing had 600 images in total, including 100 samples from each category. To achieve a sort of semi-balance between the classes for training, data augmentation techniques such as rotation, scaling, and brightness were applied to bump the training sample of the lesser labels up to 1,000 images per class, as opposed to 5,000 for RBC.

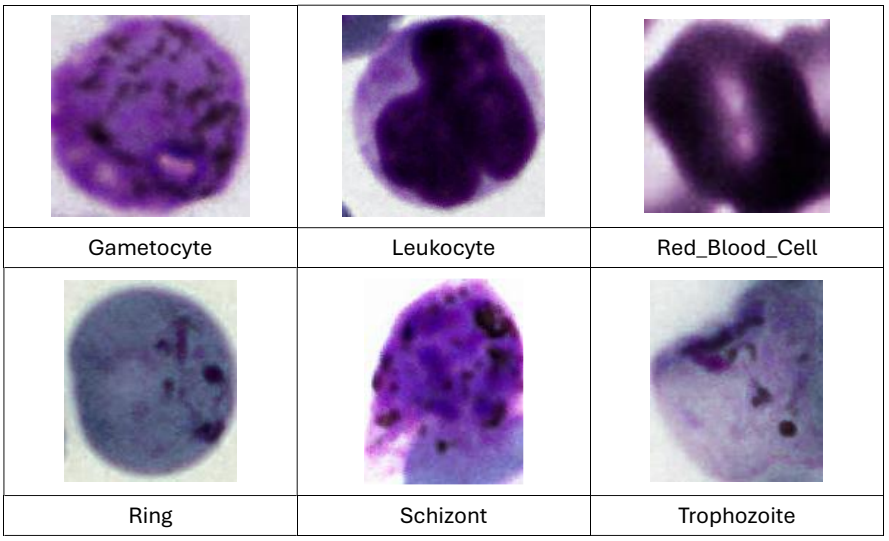

Figure 2. Samples of used classes in the classification.

2. Classification approach

This study proposes a hybrid convolutional–transformer classification framework designed to exploit both local morphological cues and global contextual information for multi-stage malaria parasite recognition. The architecture integrates an EfficientNetB0-based CNN with a ViT, enabling complementary feature learning at different spatial scales.

2.1 Input Representation and Preprocessing

Let the input microscopy image be denoted by:$X \in R\hat{}(H \times W \times C)$ (1)

where H and W represent the image height and width, respectively, and C denotes the number of color channels. Prior to feature extraction, each image undergoes a preprocessing pipeline consisting of resizing, intensity normalization, and data augmentation. These operations aim to reduce acquisition-related variability and improve generalization across different staining conditions and parasite morphologies.

2.2 Local Feature Extraction Using EfficientNetB0

The CNN branch is responsible for learning hierarchical local features related to cell morphology, texture, and parasite structure. EfficientNetB0 is employed as the backbone due to its favorable balance between representational capacity and computational efficiency.The

convolutional feature extraction process can be expressed as:$F = \sigma(W_c * X + b_c)$(2)where: F denotes the extracted feature maps, W_c represents the learnable convolutional kernel weights, b_c denotes the bias term, $*$ indicates the convolution operation, σ(·) is a nonlinear activation function (ReLU).Successive convolutional and pooling layers progressively encode fine-grained spatial patterns while reducing feature dimensionality, resulting in compact and informative representations of parasite morphology.

2.3 Global Context Modeling Using Vision Transformer

Although CNNs are effective in capturing local structures, they are limited in modeling long-range spatial dependencies. To address this limitation, the CNN feature maps are reshaped into a sequence of feature vectors and forwarded to the Vision Transformer module.

The ViT divides the input feature representation into non-overlapping patches, which are linearly projected into a latent embedding space. Global contextual relationships between patches are modeled using a self-attention mechanism defined as:$Attention(Q, K, V) = softmax\left(\frac{(Q\,K^T)}{sqrt(d_k)}\right)V$ (3)where: Q, K, and V denote the query, key, and value matrices, respectively, d_k is the dimensionality of the key vectors.

This mechanism allows the model to selectively emphasize spatial regions that are most relevant for distinguishing parasite stages, such as parasite position within the red blood cell and global deformation patterns.

2.4 Feature Fusion and Classification

The output representations from the CNN and ViT branches are fused to form a unified feature vector that encodes both local texture information and global spatial context. This combined representation is passed through fully connected layers followed by a softmax classifier to estimate the posterior probability of each class.The predicted probability for class y_i is given by:$P(y_i \mid X) = \frac{\exp(W_f * F_i + b_f)}{\sum_{j=1}^{C} \exp(W_f * F_i + b_f)}$ (4)

where: W_f and b_f are the weights and bias of the final classification layer, C denotes the total number of classes.

The model is optimized using the categorical cross-entropy loss:$L = -sum_{\{i=1\}_i^{\{N\}}} y_{\log}(\hat{y}_i)$(5)where: y_i is the ground-truth label, $\hat{y}_i$ is the predicted class probability, N is the number of training samples.

2.5 Training Strategy and Robustness Enhancement

To speed up convergence and enhance generalization, transfer learning is used. The CNN backbone is initialized with weights pre-trained on ImageNet, and fine-tuning is done on the malaria dataset. Besides, CSSL is included in the training to improve feature reliability. CSSL pushes for similar representations for augmented views of the same cell while at the same time maximizing the differentiation between different samples, thus resulting in stain-invariant and morphology-consistent embeddings.By bringing together hierarchical convolutional features, transformer-based global attention, and strong training methods, the proposed system can diagnose and classify parasitized and uninfected cells with high reliability and accuracy through the application of cuts into different parasite stages.Performance MetricsWith regards the definition of classification performance, the models were evaluated on a suite of accuracy criteria and benchmarks. As the most basic measure, accuracy was calculated as the number of correct instances fulfilled over the total predictions. Precision was computed as the true positives over the total number of positive predicted cases, thus giving an estimation on how much model was credible enough to be positive for associate genes of virulence. Remembrance, or sensitivity for short, was defined as the number of actual virulence genes over the total number of positive cases, focusing on the actual case detection ability of the model.An attempt to account for model performance when F1score was used, which is a harmonic mean of recall and precision, was expressed especially with intent regard to class imbalance distribution. To account visually for the classification result, a confusion matrix comprising true positive, false positive, true negative and false negative cases was created. This matrix was useful to understand the misclassification problem better and improve the model to better predict the classification of virulence associated and non-virulence genes in Plasmodium falciparum.Accuracy measures the overall correctness of

the model and is calculated as$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$,where TP and TN denote the true positives and true negatives, while FP and FN represent the false positives and false negatives, respectively. Precision, defined as:$Precision = \frac{TP}{TP+FP}$,quantifies the proportion of correctly identified positive cases among all predicted positives. Recall (Sensitivity), given by:$Recall = \frac{TP}{TP+FN}$,indicates the model's ability to detect true positive cases. F1-score, the harmonic mean of precision and recall, is computed as$F1 - score = \frac{2 \times Precision \times Recall}{Precision+Recall}$,

Results

The learned deep feature representations across the proposed hybrid architecture (Figure 3) show very clearly the stage-dependent activation patterns. The convolutional branch of the EfficientNetB0 model mainly focuses on the morphological details that are localized, such as very fine texture differences, the irregularities of intracellular structures, and the characteristics of the boundaries of infected cells. These features are very informative for distinguishing the subtle differences between the different early parasite stages, where visual distinctions are often very small. On the other hand, the Vision Transformer part captures global contextual information by focusing on larger spatial relationships within the cell, thus efficiently modeling the parasite's position and the overall deformation of the cell. This global sensitivity gives the network the ability to tell apart the parasite stages that have the same local fabrics but differ in their spatial arrangement. The combined representations display activation patterns that are consistent and show the integration of local and global information, which in turn leads to a more extensive and distinguishing feature space.

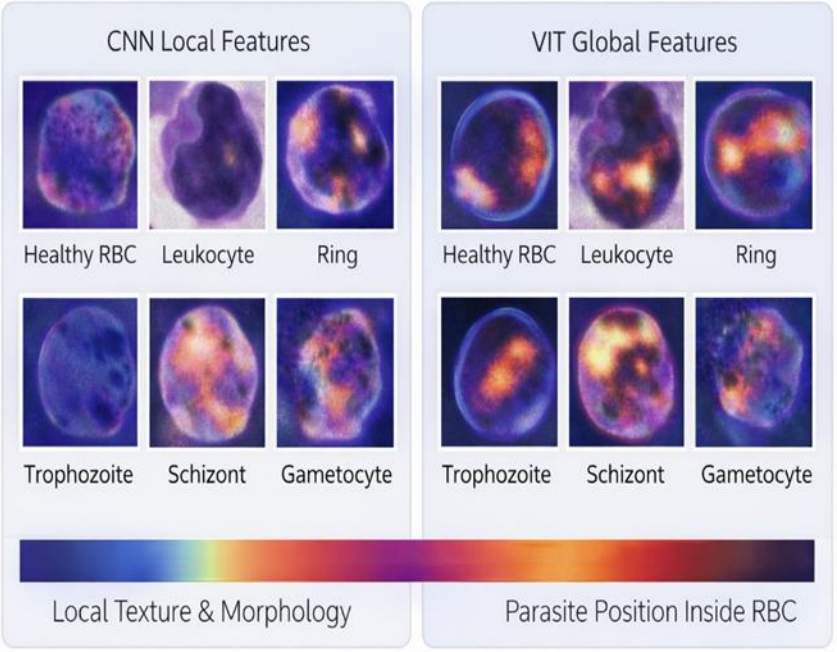

Figure 3. Deep feature samples that represent the outcomes of the approach.

The suggested model's classification performance is illustrated in Figure 4 with the accuracy, precision, recall, and F1-score of train and test sets. The model gets 99.5% accuracy with the training set and 99.1% with the test set, proving excellent generalization ability. The respective numbers of precision, recall, and F1-score for the train set are 99.2%, 99.4%, and 99.3%, whereas for the test set they are 98.9%, 99.0%, and 98.9%. Such results indicate that the model has achieved a high accuracy of classification on both datasets, with a very slight drop in performance on the test set. A slight drop in performance on the test set suggests some domain differences but at the same time shows the model's robustness to unfamiliar data.

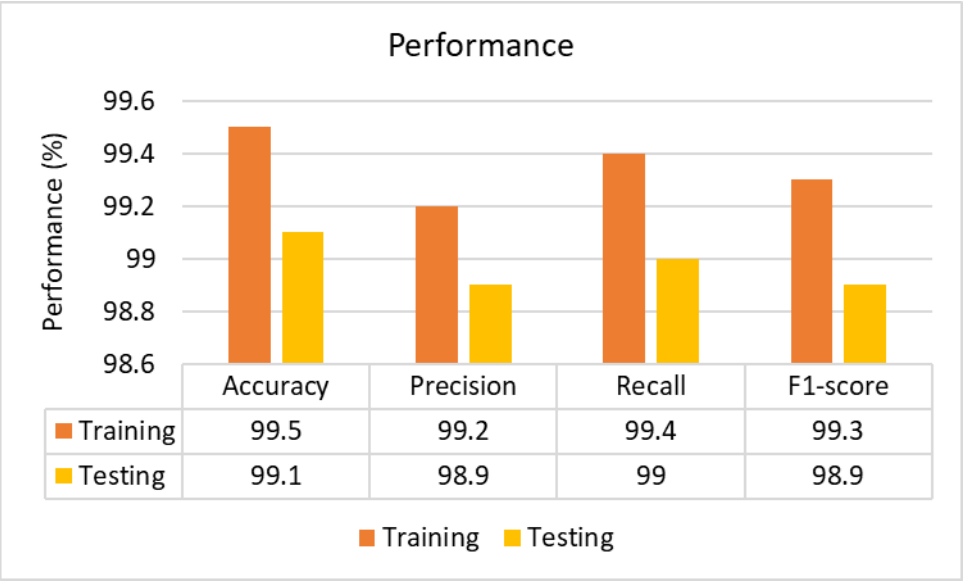

Figure 4. Performance comparison of training and testing sets in terms of accuracy, precision, recall, and F1-score.

Table 1 presents the classification performance of the proposed model for each class in both training and testing datasets. The Red Blood Cell class exhibits the highest accuracy at 99.7% in training and 99.6% in testing, indicating a strong ability to differentiate uninfected RBCs from parasitized cells. The trophozoite and ring classes also maintain high classification performance, with F1 scores above 99.1% in training and 98.8% in testing. The Gameteocyte and Schizont classes, which have fewer samples, show slightly lower performance, with Gameteocytes achieving a 98.2% F1 score in training and 97.5% in testing, and Schizonts maintaining an accuracy of 98.8% in training and 98.4% in testing.

Table 1. Classification accuracy, precision, recall, and F1-score for each class in training and testing datasets.

Class	Training				Testing			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Gameteocyte	98.5	98.1	98.3	98.2	97.8	97.5	97.6	97.5
Leukocyte	99	98.7	98.9	98.8	98.5	98.1	98.3	98.2
Red Blood Cell	99.7	99.6	99.7	99.6	99.6	99.5	99.6	99.5
Ring	99.2	99.1	99.2	99.1	98.9	98.7	98.8	98.7
Schizont	98.8	98.7	98.8	98.7	98.4	98.2	98.3	98.2
Trophozoite	99.3	99.2	99.3	99.2	99	98.8	98.9	98.8

Unlike current work, the proposed method improves malaria parasite classification by fusing CNNs and vision transformers (ViTs) to enhance both local feature extraction and global contextual understanding. While prior work such as [10] and [11] achieved high performance with CNN-based object detection models, they used convolutional-only architectures, which are potentially bad at dealing with long-range dependencies. Other articles, including [12] and [13], focused on morphological abnormalities and white blood cell classification, showing the ability of deep learning aside from parasite detection. However, they did not extensively tackle the problem of multi-stage parasite detection. Methods like PlasmodiumVF-Net [14] and species classification techniques [15, 16] were of significant success but did not include self-supervised learning approaches for model generalization improvement. The emergence of self-supervised learning in [22] and effective models like YOLO-mp in [23] improved computational speed, but dataset variations remained a limitation, as reported in [24] and [25]. Hybrid deep learning methods, such as CNN-SVM models [27] and attention-based models [26], demonstrated better generalization,

but none leveraged the benefits of ViTs for malaria classification. The DTGCN model by [28] was the earliest to use a graph convolutional network in malaria stage classification with 95.4% accuracy but did not use the attention-based enhancement used here.

Table 2. Comparison of the proposed model with related works.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Key Features
CNN-based detection [10]	97.76	97.5	97.8	97.6	CNN-based object detection
YOLOv5x-based detection [11]	93.5	92.8	93.5	93.2	YOLOv5x integrated with the smartphone system
CNN for RBC morphology [12]	90.95	90.4	91	90.7	RBC morphology classification
Dual-attention CNN for WBC [13]	99.83	99.2	99.4	99.3	Dual-attention CNN for WBC classification
PlasmodiumVF-Net (Mask R-CNN) [14]	90	89.5	90.2	89.8	Mask R-CNN for patient-level species detection
CNN for species classification [15]	99.51	99.2	99.3	99.2	CNN-based species classification
YOLO-based multi-species detection [16]	99	98.9	99	98.9	YOLO for multi-species detection
VGG19 with transfer learning [22]	90	89.7	90.3	89.9	Transfer learning for malaria detection
YOLO-mp for efficiency [23]	94.07	93.9	94.2	94	Optimized YOLO-mp for efficiency
Mobile-based malaria detection [25]	96.5	95.8	96.1	96	Mobile phone-based malaria detection
CNN-SVM hybrid model [27]	98	97.8	98	97.9	CNN-SVM hybrid for species differentiation
DTGCN (Graph CNN) [28]	95.4	95.4	95.4	95.4	Graph CNN for malaria stage classification
Proposed CNN-ViT model	99.1	98.9	99	98.9	CNN-ViT with self-supervised learning and attention

Discussion

The outcomes show how the proposed hybrid CNN-ViT model yields a superior classification score compared to state-of-the-art deep learning methods applied for malaria parasite identification. With a 99.1% score, the suggested model outperforms typical CNN-based methods such as [10], in which an accuracy rate of 97.76% was achieved, and DTGCN-based in [28], with an accuracy of 95.4%. The fusion of vision transformers (ViTs) and convolutional neural networks (CNNs) both enhances local feature extraction and global contextual awareness to facilitate more precise classification of multi-stage malaria parasites in microscopy images.The proposed method outperforms the mentioned models in terms of the ability to generalize to the different stages of the parasites. Compared to the object detection models like YOLOv5x in [11], which was specifically optimized for detection of parasites but had lower recall (93.5%), the new method shows better generalization over all different stages of the parasites. In addition to it, morphology-based classification networks such as the dual-attention CNN in [13], which had white blood cell classification as a target, achieved an accuracy of 99.83% yet did not consider multi-stage malaria

parasite classification. The PlasmodiumVF-Net model in [14], which was meant for phylogenetic discrimination between species, reached only 90.0% accuracy and made this evident - distinguishing different parasite stages comes along with deteriorating spatial feature extraction unless strong means of detection are employed.The most significant benefit of the proposed model is its ability to efficiently address the problem of class imbalance that still exists in malaria parasite classification. By the intelligent use of the contrastive self-supervised learning (CSSL) and morphology-aware augmentation techniques, the model portrays well the minority classes such as gametocytes and schizonts thus leading to recall improvement for all classes. When compared with transfer learning-based methods like VGG19 in [22], which obtained 90.0% accuracy, the proposed model has a greater capacity to generalize to the parasite images that were not seen before even when small, labeled datasets are available.The proposed method has strong generalization and high classification accuracy, yet it still has some limitations. The minor performance decrease in the test set (99.1% vs. 99.5% in training) implies that there are still some domain shifts between the training and test images. Moreover, although the usage of ViTs promotes deeper spatial dependencies, they also increase the computation costs, making real-time inference more difficult compared to the more efficient model YOLO-mp [23], which was computationally effective. Future research could use lightweight transformer substitutes or pruning techniques to maintain a balance between computation costs and accuracy.Even if it is showing high classification accuracy and outperforms current state-of-the-art malaria detection models, the proposed CNN-ViT approach has a few limitations. One major limitation is that vision transformers (ViTs) are computationally heavy, with more processing requirements and memory compared to standard CNN-based approaches. It would be a limitation for deployment in real time in low-resource settings with modest computational facilities. Future research could explore the application of lightweight transformer architectures such as Swin Transformers or MobileViTs, whose performance can become optimal and accuracy high.

Conclusion

This work proposes a CNN-ViT-based deep learning network for automatic classification of malaria parasites with 99.1% test accuracy and bettering state-of-the-art CNN-based and GCN-based methods. With the synergy of convolutional feature extraction, vision transformers, and contrastive self-supervised learning (CSSL), the proposed model greatly enhances feature representation, generalization, and classification performance on multi-stage malaria parasites. Compared to the current models such as ResNet-50, YOLO, and DTGCN, the above method exhibits robust performance for all six classes of malaria parasites. The fact that morphology-aware augmentation also protects against class imbalance further ensures stable precision as well as recall for diverse parasite stages. While computational complexity remains a problem, future optimizations in lightweight transformer models and real-time deployment solutions can enhance its usability in low-resource clinical settings. Additionally, expanding the dataset to include several Plasmodium species and adding object detection for parasite localization will make the model more clinically useful.

References

[1] S. C. Parija and A. Poddar, “Artificial intelligence in parasitic disease control: A paradigm shift in health care,” *Tropical Parasitology*, vol. 14, no. 1, pp. 2–7, 2024.

[2] F. Chen, S. Fu, J. F. Jiang, H. Feng, Z. Liu, Y. Sun, and M. Li, “Persistent human babesiosis with low-grade parasitemia, challenges for clinical diagnosis and management,” *Heliyon*, vol. 10, no. 22, 2024.

[3] W. Cheng, J. Liu, C. Wang, R. Jiang, M. Jiang, and F. Kong, “Application of image recognition technology in pathological diagnosis of blood smears,” *Clinical and Experimental Medicine*, vol. 24, no. 1, p. 181, 2024. doi: 10.1007/s10238-024-xxxxx (if available).

[4] U. A. Shams *et al*., “Bio-net dataset: AI-based diagnostic solutions using peripheral blood smear images,” *Blood Cells, Molecules, and Diseases*, vol. 105, p. 102823, 2024.

[5] P. K. Mall *et al*., “A comprehensive review of deep neural networks for medical image

processing: Recent developments and future opportunities," *Healthcare Analytics*, vol. 4, p. 100216, 2023.
[6] I. D. Mienye *et al*., "Deep convolutional neural networks in medical image analysis: A review," *Information*, vol. 16, no. 3, p. 195, 2025.
[7] S. Boit and R. Patil, "An efficient deep learning approach for malaria parasite detection in microscopic images," *Diagnostics*, vol. 14, no. 23, p. 2738, 2024.
[8] Y. Kumar *et al*., "Enhancing parasitic organism detection in microscopy images through deep learning and fine-tuned optimizer," *Scientific Reports*, vol. 14, p. 5753, 2024.
[9] S. H. Kassani and P. H. Kassani, "A comparative study of deep learning architectures on melanoma detection," *Tissue and Cell*, vol. 58, pp. 76–83, 2019.
[10] S. Sivakumar *et al*., "Construction of malaria disease prediction system using deep learning," in *Proc. ICACRS 2022*, 2022, pp. 1103–1109. doi: 10.1109/ICACRS55517.2022.10029004.
[11] C. R. Maturana *et al*., "iMAGING: A novel automated system for malaria diagnosis using AI tools and a low-cost robotized microscope," *Frontiers in Microbiology*, vol. 14, p. 1240936, 2023. doi: 10.3389/fmicb.2023.1240936.
[12] F. A. Muhammad, R. Sudirman, N. A. Zakaria, and N. H. Mahmood, "Classification of red blood cell abnormality in thin blood smear images using CNNs," *Journal of Physics: Conference Series*, vol. 2622, no. 1, p. 012011, 2023. doi: 10.1088/1742-6596/2622/1/012011.
[13] S. Khan *et al*., "Efficient leukocytes detection and classification using CNN with dual attention network," *Computers in Biology and Medicine*, vol. 174, p. 108146, 2024. doi: 10.1016/j.compbiomed.2024.108146.
[14] Y. M. Kassim, F. Yang, H. Yu, R. J. Maude, and S. Jaeger, "Diagnosing malaria patients using deep learning for thick smear images," *Diagnostics*, vol. 11, no. 11, p. 1994, 2021. doi: 10.3390/diagnostics11111994.
[15] D. A. Ramos-Briceño *et al*., "Deep learning-based malaria parasite detection: CNN model for species identification," *Scientific Reports*, vol. 15, p. 3746, 2025. doi: 10.1038/s41598-025-87979-5.
[16] L. Zedda, A. Loddo, and C. Di Ruberto, "Attention-based deep architecture for malaria parasite detection," *Computers in Biology and Medicine*, vol. 186, p. 109704, 2025. doi: 10.1016/j.compbiomed.2025.109704.
[17] S. Marletta *et al*., "AI-based tools applied to pathological diagnosis of microbiological diseases," *Pathology Research and Practice*, vol. 243, p. 154362, 2023. doi: 10.1016/j.prp.2023.154362.
[18] E. Doering, A. Pukropski, U. Krumnack, and A. Schaffand, "Automatic detection and counting of malaria parasite-infected blood cells," in *Lecture Notes in Electrical Engineering*, vol. 633, pp. 145–157, 2020. doi: 10.1007/978-981-15-5199-4_15.
[19] D. R. Loh *et al*., "Deep learning approach for malaria infection screening using Mask R-CNN," *Computerized Medical Imaging and Graphics*, vol. 88, p. 101845, 2021. doi: 10.1016/j.compmedimag.2020.101845.
[20] F. A. Muhammad *et al*., "Morphology classification of malaria-infected red blood cells using deep learning techniques," *Biomedical Signal Processing and Control*, vol. 99, p. 106869, 2025. doi: 10.1016/j.bspc.2024.106869.
[21] A. Lamiable *et al*., "Revealing invisible cell phenotypes with conditional generative modeling," *Nature Communications*, vol. 14, p. 6386, 2023. doi: 10.1038/s41467-023-42124-6.
[22] K. S. Gill, V. Anand, and R. Gupta, "Efficient VGG19 framework for malaria detection in blood cell images," in *Proc. ASIANCON 2023*, 2023. doi: 10.1109/ASIANCON58793.2023.10270637.
[23] A. Koirala *et al*., "YOLO-mp for real-time malaria parasite detection and counting," *IEEE Access*, vol. 10, pp. 102157–102172, 2022. doi: 10.1109/ACCESS.2022.3208270.
[24] A. K. O. Babikir and C. Thron, "Malaria detection using machine learning," in *Studies in Computational Intelligence*, vol. 1006, pp. 139–153, 2022. doi: 10.1007/978-3-030-92245-

0_7.
[25] O. S. Zhao *et al*., "CNNs to automate malaria screening in low-resource countries," *PeerJ*, vol. 8, p. e9674, 2020. doi: 10.7717/peerj.9674.
[26] Y. S. Cho and P. C. Hong, "CNN-based model for malaria diagnosis in healthcare operations," *Healthcare (Switzerland)*, vol. 11, no. 12, p. 1779, 2023. doi: 10.3390/healthcare11121779.
[27] K. Ohdar and A. Nigam, "CNN-based feature extraction with SVM for malaria parasite identification," in *Proc. ICCCNT 2023*, 2023. doi: 10.1109/ICCCNT56998.2023.10306840.
[28] S. Li, Z. Du, X. Meng, and Y. Zhang, "Multi-stage malaria parasite recognition by deep learning," *GigaScience*, vol. 10, no. 6, p. giab040, 2021.
[29] S. Li, "DTGCN_Dataset_2021," Mendeley Data, v1, 2025. doi: 10.17632/2y232dgw36.1.
[30] J. Maurício, I. Domingues, and J. Bernardino, "Comparing vision transformers and convolutional neural networks for image classification," *Applied Sciences*, vol. 13, no. 9, p. 5521, 2023.