PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol. 05, No.2, September 2026
Pages 361-372

Malaria Parasite Detection Using Deep Learning: A Comparative Study of Six Architectures

Muhamad Syukron[1*], Nadia Putri Dira[2], Nabila Sinta Kusumadewi[3]

Department of Statistics, Faculty of Sains and Data Analyitcs,
Institut Teknologi Sepuluh Nopember, Surabaya, 60111, East Java, Indonesia

E-mail Correspondence Author: *muhamad.syukron@its.ac.id

Abstract

Malaria remains a major global health challenge, particularly in resource-limited regions where reliable diagnostic tools are scarce. Traditional thick blood smear microscopy, while accurate, is time-consuming and requires expert interpretation. This study employed Artificial Intelligence as a pre-diagnostic tool to mitigate the impact of malaria by evaluating six deep learning architectures for classifying malaria-infected cell images: Vanilla CNN, AlexNet, ViT, Swin Transformer, ConvNeXt, and MaxViT. The dataset consisted of 1,883 Giemsa-stained blood smear samples collected from patients at Chittagong Medical College Hospital, Bangladesh, with expert annotations from the Mahidol Oxford Tropical Medicine Research Unit. All models were trained for 100 epochs using the Adam optimizer, a cosine annealing learning rate scheduler, early stopping, and a batch size of 32 samples. Among the tested models, MaxViT yielded the highest accuracy of 99.67%, followed by ConvNeXt with 99.61% and Swin Transformer with 98.79%, while ViT reached 80.03%, AlexNet 77.83%, and Vanilla CNN 50.06%. These findings demonstrate that more advanced architectures integrating both convolutional and attention mechanisms yield superior performance in malaria cell classification. With their high accuracy, cost-effectiveness, and computational efficiency, these architectures show strong potential as reliable pre-diagnostic tools for malaria detection in resource-constrained settings.

Keywords: Malaria detection, computer vision, CNN based models, transformer based models, hybrid deep learning.

: https://doi.org/10.30598/parameterv5i2pp361-372

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright © 2026 The Author(s).
Research Ariticle ***Open Acces***
Email: jurnalparameter@gmail.com
Homepage: https://ojs3.unpatti.ac.id/index.php/parameter

Submitted: April 2026
Accepted: September 2026

1. INTRODUCTION

Malaria is an endemic disease that affects 84 countries worldwide and caused approximately 247 million cases and 619,000 deaths in 2021 [1], [2]. The overwhelming majority of these cases occur in sub-Saharan Africa, which bears the heaviest malaria burden globally and continues to be the most severely affected region [3]. Children under the age of five represent the most vulnerable population group, accounting for approximately 80% of malaria-related deaths in the region, underscoring their heightened susceptibility [4]. This significant burden highlights the urgent need for accurate and early malaria diagnosis to improve health outcomes and reduce mortality rates in vulnerable populations [5], [6].

In conventional clinical practice, malaria detection primarily relies on the manual examination of blood samples under a microscope, in which experts individually identify parasite patterns indicative of infection. Although this remains the gold standard, the method is time-consuming, labor-intensive, and requires highly trained personnel, limiting its availability in many endemic regions [7], [8]. Despite significant improvements in diagnostic techniques over recent years, reductions in malaria mortality have been slow, largely due to insufficient healthcare infrastructure and a scarcity of skilled microscopists in resource-limited settings [9]. This gap highlights the urgent need for alternative diagnostic approaches that can maintain diagnostic accuracy while overcoming these operational constraints. To address these limitations, automated systems leveraging artificial intelligence (AI) have been proposed for malaria parasite detection in thick blood smear images [10]. AI-driven diagnostic models offer considerable advantages, including rapid analysis and accuracy comparable to that of expert microscopists [11]. Furthermore, the deployment of such models via mobile or web-based platforms facilitates broader accessibility and real-time detection in remote and underserved areas. The integration of AI into point-of-care testing thus holds transformative potential to democratize malaria diagnosis, enhance early detection and treatment, and contribute to global malaria control efforts [12], [13], [14].

Previous studies have attempted to develop automated systems for malaria detection. However, most of these studies relied on earlier-generation machine learning models [15], [16], [17], [18]. For instance, Rosado et al. [19] developed a system for malaria parasite detection using a Support Vector Machine (SVM) [20], achieving a sensitivity of 80.5%. SVM's strength lies in its ability to handle high-dimensional data and find optimal hyperplanes for separation; however, it often requires manual feature engineering, which can be time-consuming and less effective for complex image patterns in malaria diagnostics. As deep learning has gained prominence, this algorithm has been used for malaria detection [21], [22], [23], [24], [25]. Torres et al. [26] evaluated CNNs [27] for malaria parasite detection using digital microscopy data. Their model achieved a sensitivity and specificity of 72% and 85%, respectively. Yang et al. [28] also developed a CNN to classify blood sample images as parasitized or non-parasitized. They used Iterative Global Minimum Screening (IGMS) for parasite candidate selection and achieved 97.26% accuracy. Building on earlier studies, hybrid approaches that combine CNNs with traditional classifiers, such as SVM, have also been explored. Mehanian et al. [29] proposed a hybrid method for malaria parasite detection and achieved a sensitivity of 91.6%. Although this approach outperformed a standalone SVM by utilizing CNN-based feature extraction, both models belong to earlier generations of machine and deep learning methods. These models were introduced over 25 years ago, which limits their suitability for modern high-performance applications. Consequently,

recent studies favor either single deep learning models or combinations of multiple deep learning architectures to achieve higher accuracy and better computational efficiencies.

CNN-based architectures are particularly well-suited for microscopy image analysis due to their ability to automatically extract hierarchical spatial features, such as cell morphology, texture, and color variations, which are critical for distinguishing parasitized from healthy cells. Transformer-based models, on the other hand, leverage self-attention mechanisms to capture long-range dependencies across image regions, enabling a more holistic understanding of cellular context that local convolutional filters may miss. Hybrid architectures that integrate both convolutional and attention-based operations offer a complementary approach, combining the local feature extraction strength of CNNs with the global contextual modeling capability of Transformers, making them particularly promising for complex microscopy image classification tasks. Despite the growing body of research on deep learning for malaria detection, a systematic and rigorous comparison of architectures spanning conventional CNNs, modern Transformer-based models, and hybrid designs on a standardized malaria microscopy dataset remains lacking. Most existing studies evaluate only a single model or a narrow set of architectures, making it difficult to draw meaningful conclusions about the relative strengths and limitations of each approach under consistent experimental conditions. Furthermore, the extent to which architectural advances translate into tangible performance gains for malaria cell classification has not been thoroughly investigated.

To address this gap, this study presents a comprehensive benchmarking of six deep learning architectures for malaria parasite detection on a Giemsa-stained blood smear dataset collected from Chittagong Medical College Hospital, Bangladesh. We evaluate a conventional CNN as a baseline alongside AlexNet **[30]**, ViT **[31]**, Swin Transformer **[32]**, ConvNeXt **[33]**, and MaxViT **[34]**, all trained under identical experimental conditions using three-fold cross-validation. The primary contribution of this work is a systematic performance comparison that provides empirical insights into the suitability of each architecture for malaria cell classification, offering practical guidance for the development of AI-based diagnostic tools in resource-constrained healthcare settings.

2. METHOD

A vanilla CNN serves as the baseline model in this study, consisting of convolutional layers for local feature extraction, pooling layers for spatial dimensionality reduction, and fully connected layers for classification (**Figure 1**). Convolutional layers apply learned filters to detect low-level patterns such as edges, textures, and cell morphology, which are diagnostically relevant in blood smear images. This architecture provides a reference point for assessing the performance gains offered by more advanced designs.

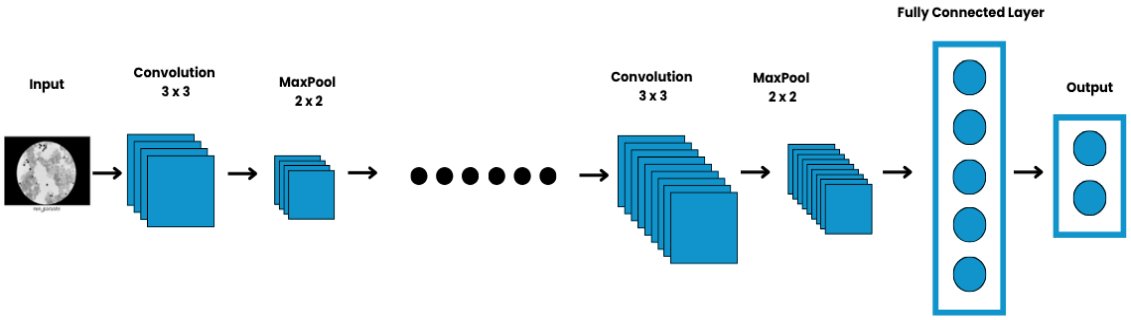


Figure 1. CNN Architecture

AlexNet is a foundational deep learning architecture consisting of five convolutional layers and three fully connected layers (**Figure 2**). It introduced key components such as ReLU activation, dropout regularization, and overlapping max pooling, which were influential in advancing image classification. While no longer considered a modern architecture, AlexNet is included here as a historically significant benchmark to contextualize the performance of more recent models.

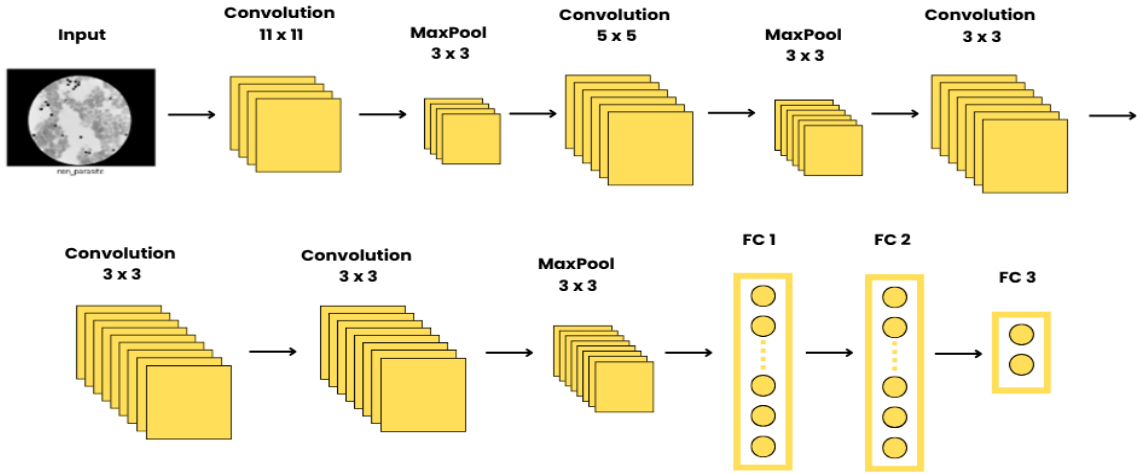


Figure 2. AlexNet Architecture

The Vision Transformer (ViT) divides the input image into fixed-size 16×16 patches, each treated as a token in a sequence (**Figure 3**). A learnable [CLS] token is prepended to represent the global image context, and positional embeddings are added to preserve spatial information. The token sequence is processed through multiple multi-head self-attention layers, enabling all patches to interact with one another regardless of their spatial distance. This global receptive field allows ViT to capture relationships between spatially separated cellular structures, which local convolutional operations may fail to model effectively.

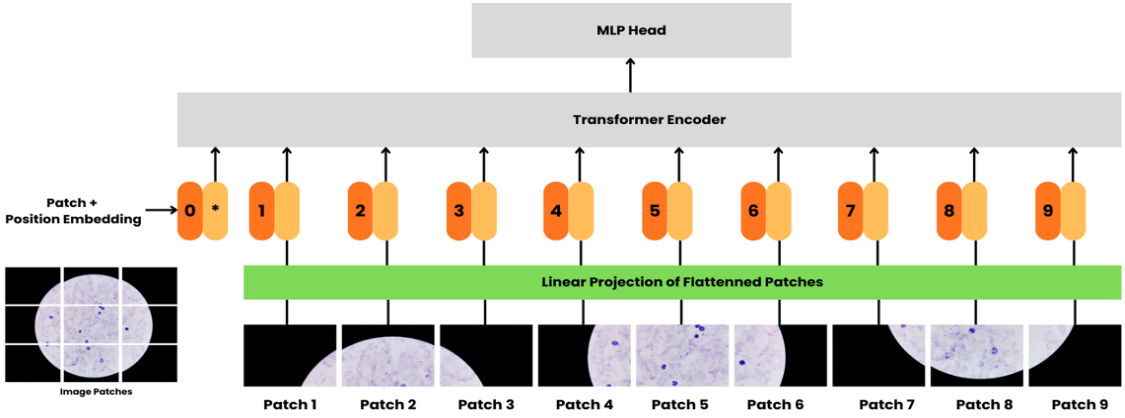


Figure 3. ViT Architecture

The Swin Transformer addresses the computational cost of global self-attention by restricting attention computation to local non-overlapping windows of size 7×7 (**Figure 4**). To enable cross-window information exchange, it alternates between regular window attention (W-MSA) and shifted window attention (SW-MSA), where window positions are offset by a fixed number of pixels between successive blocks. This hierarchical design progressively reduces spatial resolution across four stages, building representations from fine-grained local features to broader semantic context, making it computationally more efficient than ViT while retaining strong global modeling capability.

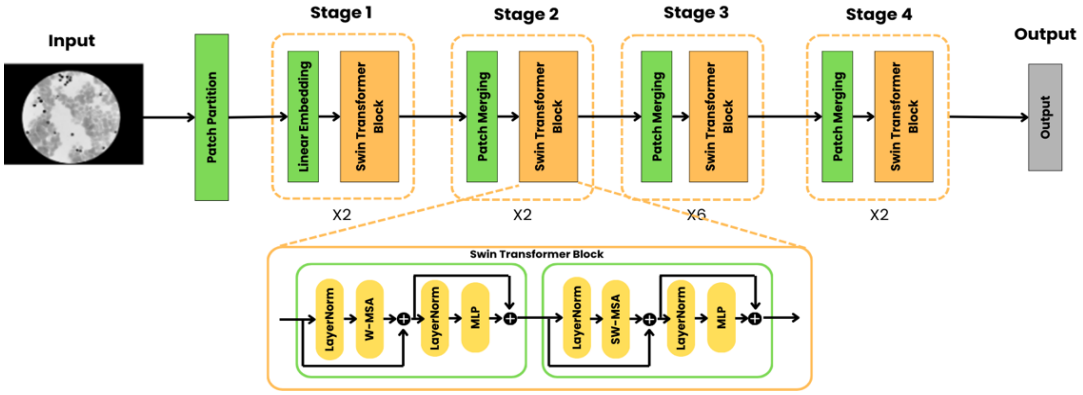


Figure 4. Swin Transformer Architecture

ConvNeXt is a modern convolutional architecture that incorporates several design principles inspired by Transformer architectures, including large 7×7 depthwise convolutions, Layer Normalization, and an inverted bottleneck block structure (**Figure 5**). These choices allow ConvNeXt to achieve a substantially larger receptive field than conventional CNNs, enabling it to capture broader cellular context in blood smear images without relying on self-attention mechanisms. The architecture consists of four stages with 3, 3, 9, and 3 ConvNeXt blocks respectively, where deeper stages capture increasingly abstract features relevant to parasite classification.

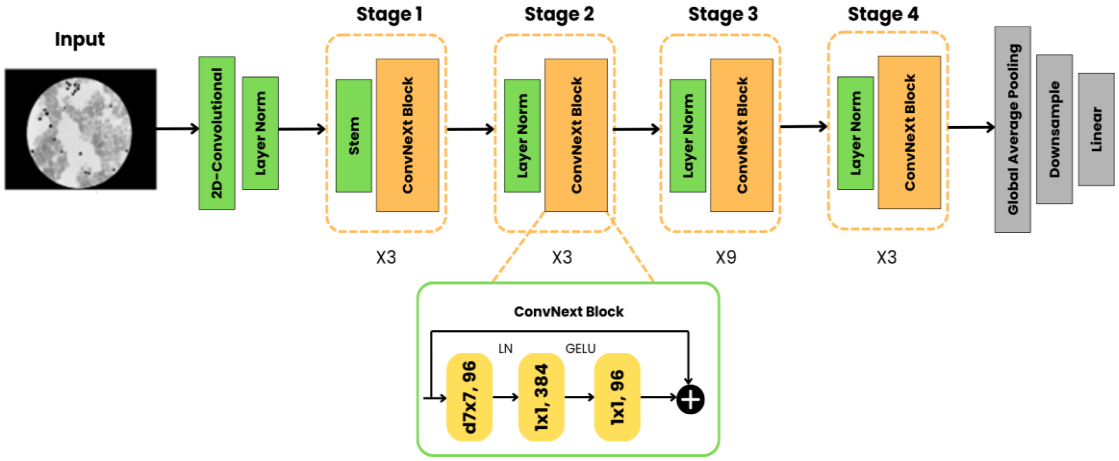


Figure 5. ConvNeXt Architecture

MaxViT integrates both convolutional and self-attention mechanisms within a single block (**Figure 6**). Each MaxViT block consists of three sequential components: MBConv, Block Attention, and Grid Attention. The MBConv component performs local feature extraction using depthwise convolution augmented by a Squeeze-and-Excitation mechanism for channel recalibration, providing sensitivity to fine-grained morphological details such as parasite ring forms or chromatin dots. Block Attention then applies self-attention within non-overlapping local windows to capture short-range contextual relationships, while Grid Attention computes self-attention over dilated global grids, enabling the model to relate distant regions of the blood smear without the quadratic cost of full self-attention. This multi-scale design allows MaxViT to simultaneously exploit local cellular features and global slide-level context, which is particularly relevant for distinguishing parasitized from healthy cells in heterogeneous blood smear images.

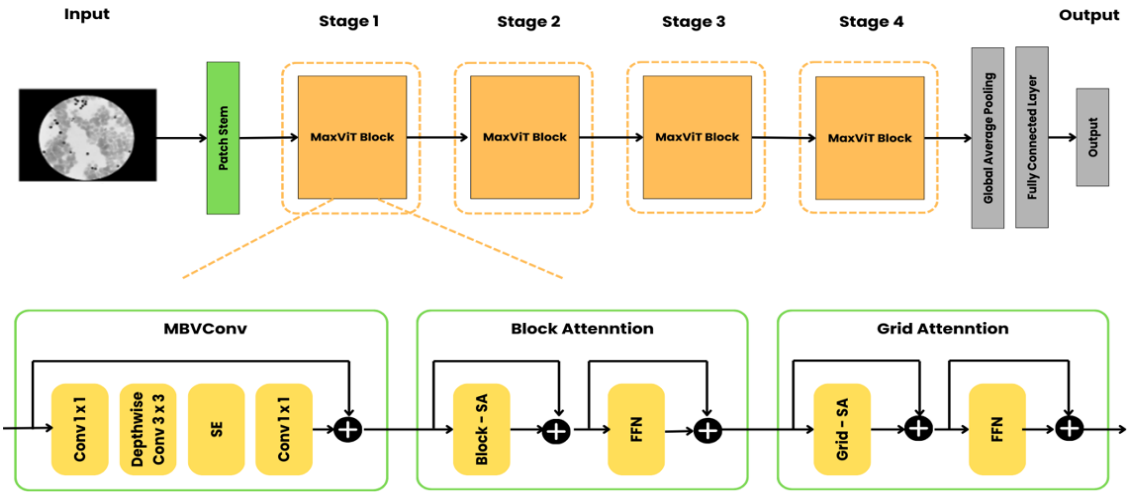


Figure 6. MaxViT Architecture

3. RESULTS AND DISCUSSION

3.1. Dataset Preparation

The dataset used in this study consists of Giemsa-stained thick blood smear images collected from 150 malaria patients and 50 non-malaria patients at Chittagong Medical College Hospital, Bangladesh **[28]**. The images were captured using a smartphone

camera and manually annotated by experts at the Mahidol–Oxford Tropical Medicine Research Unit (MORU) in Bangkok, Thailand. The dataset was anonymized and archived by the National Library of Medicine and contains two classes: malaria and non-malaria, with 1,883 malaria images and 1,141 non-malaria images. Sample images from each class are shown in **Figure 7**. Malaria images contain parasites, whereas non-malaria images do not contain parasites.

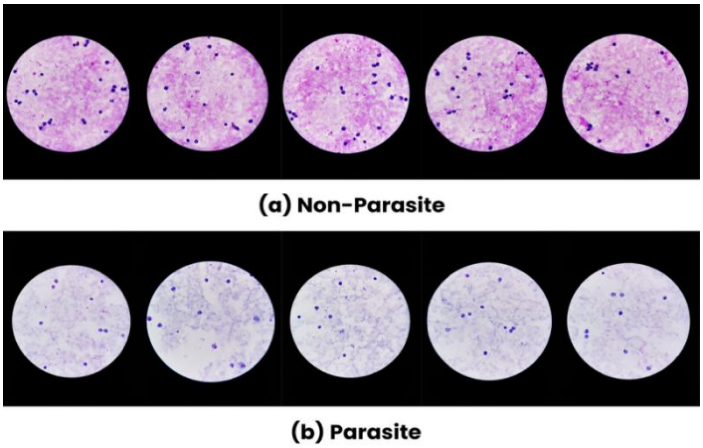


Figure 7. Example of Thick Blood Smear Image

A train–test split was performed with an 80:20 ratio, resulting in 1,506 malaria and 912 non-malaria images for training, and 377 malaria and 229 non-malaria images for testing, as summarized in **Table 1**.

Table 1. Class Distribution of Dataset Before Undersampling

	Malaria	Non-Malaria	Total
Train	1,506	912	2,418
Test	377	229	606
Total	1,883	1,141	3,024

The training set exhibited a class imbalance between malaria and non-malaria images. To address this, random undersampling was applied to the majority malaria class to match the non-malaria class, resulting in a balanced training set of 912 images per class, for a total of 1,824 training images. The resulting class distribution after undersampling is shown in **Table 2**.

Table 2. Class Distribution of Dataset After Undersampling

	Malaria	Non-Malaria	Total
Train	912	912	1,824
Test	377	229	606
Total	1,289	1,141	2,430

Three-fold cross-validation was then applied to this balanced training set, with each fold containing 1,216 training images (608 per class) and 608 validation images (304 per class), as shown in **Table 3**.

Table 3. Data Split per Fold in Three-Fold Cross-Validation

	Malaria	Non-Malaria	Total
Train	608	608	1,216
Validation	304	304	608
Total	912	912	1,824

Before training, all images were standardized to 224×224 pixels using TensorFlow, with pixel values normalized to the range [0, 1]. Data augmentation was applied to the training set, including vertical flips, 90° rotations, brightness adjustments, and Gaussian blur to enhance generalization. The test set underwent only resizing and normalization to maintain evaluation consistency.

3.2. Experiment Result

This study evaluates the performance of six deep learning architectures for malaria parasite detection: Vanilla CNN, AlexNet, ViT, Swin Transformer, ConvNeXt, and MaxViT. Each model was trained for a maximum of 100 epochs, with early stopping applied based on the average validation loss across three folds, halting training if no improvement was observed for 3 consecutive epochs. All metrics reported are computed with malaria as the positive class using macro-averaging to account for the class distribution in the test set. All models stopped considerably earlier than the 100-epoch maximum, with the earliest convergence observed in Vanilla CNN and Swin Transformer at epoch 12, followed by AlexNet and ConvNeXt at epoch 13, MaxViT at epoch 15, and ViT at epoch 18. This early convergence is primarily attributed to the use of pre-trained weights through transfer learning, which initializes the models with rich feature representations learned from large-scale datasets such as ImageNet. As a result, the models require far fewer gradient updates to adapt to the malaria classification task compared to training from scratch. The cosine annealing learning rate scheduler further facilitates rapid convergence by smoothly reducing the learning rate, preventing oscillation around the optimum. The small and balanced nature of the training set after undersampling additionally accelerates convergence, as the models reach a stable optimum without requiring extensive data iterations. The classification performance on the validation set across three folds is reported in **Table 4,** with results expressed as mean (variance) to reflect the consistency of each model across folds.

Table 4. Model Comparisson Based on Val Data

Model	Precision	Recall	F1-Score	Accuracy
Vanilla CNN	49.56% (0.0984%)	49.01% (0.6468%)	49.27% (0.1924%)	49.56% (0.0968%)
AlexNet	77.83% (0.0131%)	77.74% (0.8954%)	77.80% (0.1609%)	77.80% (0.0704%)
ViT	79.85% (0.0563%)	80.81% (0.8874%)	80.32% (0.1198%)	80.21% (0.0431%)
Swin Transformer	99.02% (0.0704%)	99.23% (0.0242%)	99.12% (0.0061%)	99.12% (0.0064%)
ConvNeXt	99.67% (0.0726%)	99.78% (0.0242%)	99.73% (0.0242%)	99.73% (0.0242%)
MaxViT	99.67% (0.0717%)	99.89% (0.0240%)	99.78% (0.0420%)	99.78% (0.0421%)

The classification performance on the held-out test set is reported in **Table 5**, again expressed as mean (variance) across three folds.

Table 5. Model Comparisson Based on Test Data

Model	Precision	Recall	F1-Score	Accuracy
Vanilla CNN	62.30% (0.6635%)	49.95% (0.4370%)	55.45% (0.5324%)	50.06% (0.6661%)
AlexNet	84.94% (0.1568%)	78.25% (0.4214%)	81.45% (0.0275%)	77.83% (0.0064%)
ViT	86.65% (0.3931%)	80.65% (0.3931%)	83.32% (0.1710%)	80.03% (0.0726%)
Swin Transformer	98.68% (0.0434%)	99.38% (0.0162%)	99.00% (0.0029%)	98.79% (0.0064%)
ConvNeXt	99.47% (0.0468%)	99.91% (0.0162%)	99.69% (0.0278%)	99.61% (0.0416%)
MaxViT	99.56% (0.0610%)	99.91% (0.0156%)	99.74% (0.0116%)	99.67% (0.0182%)

The Vanilla CNN achieved the weakest performance, with a test accuracy of 50.06% and an F1-score of 55.45%, performing at near-random levels. This indicates that its shallow architecture lacks the representational capacity to distinguish subtle morphological differences between parasitized and healthy cells. The high variance across folds further reflects its instability and sensitivity to data partitioning. AlexNet demonstrated a clear improvement over Vanilla CNN, achieving a test accuracy of 77.83% and an F1-score of 81.45%. While AlexNet is no longer considered a modern architecture, its deeper convolutional structure enables more discriminative feature extraction, resulting in more stable predictions as evidenced by its lower variance. ViT achieved comparable performance to AlexNet, with a test accuracy of 80.03% and F1-score of 83.32%, suggesting that pure attention-based models without convolutional inductive biases may require larger datasets to fully exploit their global modeling capability. The Swin Transformer substantially outperformed both AlexNet and ViT, achieving a test accuracy of 98.79% and F1-score of 99.00%. Its hierarchical window-based attention mechanism enables effective modeling of both local and global features, which is particularly advantageous for blood smear image analysis where parasite morphology varies across spatial scales. The low variance across folds confirms the robustness of this result. ConvNeXt and MaxViT achieved the highest overall performance, with test accuracies of 99.61% and 99.67% and F1-scores of 99.69% and 99.74%, respectively. The consistently low variance across all metrics for both models indicates stable and reliable generalization to unseen data. The strong performance of ConvNeXt demonstrates that a well-designed modern convolutional architecture with a large receptive field can be highly competitive without relying on self-attention mechanisms. MaxViT's marginal advantage suggests that the explicit integration of local and global attention within each block provides additional discriminative benefit for this task. Beyond classification performance, practical deployment considerations are also relevant. Vanilla CNN and AlexNet are computationally lightweight and fast to infer, but their low accuracy renders them unsuitable for clinical use. ViT, while competitive in accuracy, involves substantial computational overhead due to its global self-attention mechanism, which may limit deployment on resource-constrained devices. The Swin Transformer offers a strong balance between accuracy and computational efficiency through its local

windowed attention design. ConvNeXt benefits from pure convolutional operations, which are highly optimized on most hardware and thus easier to deploy in low-resource settings. MaxViT, while achieving the best accuracy, involves greater architectural complexity due to its multi-scale attention blocks, which may require more careful optimization for edge deployment. Overall, these findings suggest that ConvNeXt and MaxViT are the most promising architectures for malaria parasite detection, offering high and consistent accuracy across folds. However, as evaluation was conducted on a single dataset, external validation on independent clinical populations from diverse geographic settings is necessary before these models can be considered for real-world pre-diagnostic applications.

4. CONCLUSION

This study developed an image classification framework for malaria parasite detection in thick blood smear images and evaluated six deep learning architectures: Vanilla CNN, AlexNet, ViT, Swin Transformer, ConvNeXt, and MaxViT. The results demonstrate clear and consistent performance differences across all models, with more advanced architectures providing substantially more reliable outcomes. Among the evaluated models, MaxViT achieved the highest test accuracy of 99.67% with an F1-score of 99.74%, followed closely by ConvNeXt with 99.61% accuracy and 99.69% F1-score, and Swin Transformer with 98.79% accuracy and 99.00% F1-score. The low variance across three-fold cross-validation for these models indicates stable and consistent generalization to unseen data. In contrast, Vanilla CNN performed at near-random levels with a test accuracy of 50.06%, highlighting the limitations of shallow architectures for complex medical image analysis. AlexNet and ViT achieved moderate performance, with test accuracies of 77.83% and 80.03% respectively, suggesting that both architectural depth and the availability of sufficient training data are critical factors in determining classification performance. These findings suggest that architectures capable of modeling both local and global spatial context, such as MaxViT, ConvNeXt, and Swin Transformer, offer superior representational capacity for malaria parasite detection. However, as all evaluations were conducted on a single dataset from a specific clinical setting, these results should be interpreted as indicative rather than conclusive. External validation on independent clinical populations from diverse geographic settings is necessary before any of these models can be considered for real-world pre-diagnostic applications. Future work should focus on validating the proposed framework on larger and more diverse datasets, as well as exploring deployment on edge or mobile platforms to enhance practical applicability in resource-constrained healthcare environments.

REFERENCES

[1] M. Balmith, C. Basson, and S. J. Brand, “The Malaria Burden: A South African Perspective,” J. Trop. Med., vol. 2024, pp. 1–17, Feb. 2024, doi: 10.1155/2024/6619010.

[2] M. González-Sanz, P. Berzosa, and F. F. Norman, “Updates on Malaria Epidemiology and Prevention Strategies,” Curr. Infect. Dis. Rep., vol. 25, no. 7, pp. 131–139, Jul. 2023, doi: 10.1007/s11908-023-00805-9.

[3] Q. Yu et al., “Spatial spillovers of violent conflict amplify the impacts of climate variability on malaria risk in sub-Saharan Africa,” Proceedings of the National Academy of Sciences, vol. 121, no. 15, Apr. 2024, doi: 10.1073/pnas.2309087121.

[4] United Nations Children's Fund (UNICEF), "Nearly Every Minute, A Child Under 5 Dies of Malaria." Accessed: Jan. 15, 2026. [Online]. Available: https://data.unicef.org/topic/child-health/malaria/

[5] P. E. Obasohan, S. J. Walters, R. Jacques, and K. Khatab, "A Scoping Review of Selected Studies on Predictor Variables Associated with the Malaria Status among Children under Five Years in Sub-Saharan Africa," Int. J. Environ. Res. Public Health, vol. 18, no. 4, p. 2119, Feb. 2021, doi: 10.3390/ijerph18042119.

[6] J. O. Sarfo et al., "Malaria amongst children under five in sub-Saharan Africa: a scoping review of prevalence, risk factors and preventive interventions," Eur. J. Med. Res., vol. 28, no. 1, p. 80, Feb. 2023, doi: 10.1186/s40001-023-01046-1.

[7] F. Grignaffini, P. Simeoni, A. Alisi, and F. Frezza, "Computer-Aided Diagnosis Systems for Automatic Malaria Parasite Detection and Classification: A Systematic Review," Electronics (Basel)., vol. 13, no. 16, p. 3174, Aug. 2024, doi: 10.3390/electronics13163174.

[8] C. R. Maturana et al., "Advances and challenges in automated malaria diagnosis using digital microscopy imaging with artificial intelligence tools: A review," Front. Microbiol., vol. 13, Nov. 2022, doi: 10.3389/fmicb.2022.1006659.

[9] Md. S. Rahman and Md. A. B. Shiddik, "Unraveling global malaria incidence and mortality using machine learning and artificial intelligence–driven spatial analysis," Sci. Rep., vol. 15, no. 1, p. 28334, Aug. 2025, doi: 10.1038/s41598-025-12872-0.

[10] D. A. Ramos-Briceño, A. Flammia-D'Aleo, G. Fernández-López, F. S. Carrión-Nessi, and D. A. Forero-Peña, "Deep learning-based malaria parasite detection: convolutional neural networks model for accurate species identification of Plasmodium falciparum and Plasmodium vivax," Sci. Rep., vol. 15, no. 1, p. 3746, Jan. 2025, doi: 10.1038/s41598-025-87979-5.

[11] A. Salam et al., "Embedded System-Based Malaria Detection From Blood Smear Images Using Lightweight Deep Learning Model," Int. J. Imaging Syst. Technol., vol. 34, no. 6, Nov. 2024, doi: 10.1002/ima.23205.

[12] T. S. Pillay, A. I. Khan, and S. Yenice, "Artificial intelligence (AI) in point-of-care testing," Clinica Chimica Acta, vol. 574, p. 120341, Jun. 2025, doi: 10.1016/j.cca.2025.120341.

[13] R. Yuriy, O. Tatarina, V. Kaminskyy, T. Silina, and L. Bashkirova, "Modern Methods and Prospects for Using Artificial Intelligence in Disease Diagnostics: A Narrative Review," Futurity Medicine, vol. 3, no. 4, Jul. 2024, doi: 10.57125/FEM.2024.12.30.02.

[14] S. S. and S. H. D., "Malaria cell identification using improved machine learning and modified deep learning architecture," Indonesian Journal of Electrical Engineering and Computer Science, vol. 34, no. 3, p. 2078, Jun. 2024, doi: 10.11591/ijeecs.v34.i3.pp2078-2086.

[15] M. Mariki, E. Mkoba, and N. Mduma, "Combining Clinical Symptoms and Patient Features for Malaria Diagnosis: Machine Learning Approach," Applied Artificial Intelligence, vol. 36, no. 1, Dec. 2022, doi: 10.1080/08839514.2022.2031826.

[16] L. Balerdi-Sarasola et al., "MALrisk: a machine-learning–based tool to predict imported malaria in returned travellers with fever," J. Travel Med., vol. 31, no. 8, Dec. 2024, doi: 10.1093/jtm/taae054.

[17] N. Rismayanti, "Segmentation and Feature Extraction for Malaria Detection in Blood Smears," International Journal of Artificial Intelligence in Medical Issues, vol. 2, no. 1, pp. 18–29, May 2024, doi: 10.56705/ijaimi.v2i1.138.

[18] M. Golumbeanu et al., "Leveraging mathematical models of disease dynamics and machine learning to improve development of novel malaria interventions," Infect. Dis. Poverty, vol. 11, no. 1, p. 61, Jun. 2022, doi: 10.1186/s40249-022-00981-1.
[19] 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 Comput. Sci., vol. 90, pp. 138–144, 2016, doi: 10.1016/j.procs.2016.07.024.
[20] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
[21] D. Sukumarran et al., "An optimised YOLOv4 deep learning model for efficient malarial cell detection in thin blood smear images," Parasit. Vectors, vol. 17, no. 1, p. 188, Apr. 2024, doi: 10.1186/s13071-024-06215-7.
[22] F. Yang, H. Yu, K. Silamut, R. J. Maude, S. Jaeger, and S. Antani, "Smartphone-Supported Malaria Diagnosis Based on Deep Learning," 2019, pp. 73–80. doi: 10.1007/978-3-030-32692-0_9.
[23] Y. S. Cho and P. C. Hong, "Applying Machine Learning to Healthcare Operations Management: CNN-Based Model for Malaria Diagnosis," Healthcare, vol. 11, no. 12, p. 1779, Jun. 2023, doi: 10.3390/healthcare11121779.
[24] A. K. Dutta et al., "Barnacles Mating Optimizer with Deep Transfer Learning Enabled Biomedical Malaria Parasite Detection and Classification," Comput. Intell. Neurosci., vol. 2022, pp. 1–12, Jun. 2022, doi: 10.1155/2022/7776319.
[25] M. Masud et al., "Leveraging Deep Learning Techniques for Malaria Parasite Detection Using Mobile Application," Wirel. Commun. Mob. Comput., vol. 2020, pp. 1–15, Jul. 2020, doi: 10.1155/2020/8895429.
[26] K. Torres et al., "Automated microscopy for routine malaria diagnosis: a field comparison on Giemsa-stained blood films in Peru," Malar. J., vol. 17, no. 1, p. 339, Dec. 2018, doi: 10.1186/s12936-018-2493-0.
[27] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.
[28] F. Yang et al., "Deep Learning for Smartphone-Based Malaria Parasite Detection in Thick Blood Smears," IEEE J. Biomed. Health Inform., vol. 24, no. 5, pp. 1427–1438, May 2020, doi: 10.1109/JBHI.2019.2939121.
[29] C. Mehanian et al., "Computer-Automated Malaria Diagnosis and Quantitation Using Convolutional Neural Networks," in 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), IEEE, Oct. 2017, pp. 116–125. doi: 10.1109/ICCVW.2017.22.
[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[31] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Jun. 2021.
[32] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," Aug. 2021.
[33] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," Mar. 2022.
[34] Z. Tu et al., "MaxViT: Multi-Axis Vision Transformer," Sep. 2022.