Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00349-8

Original Article

Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models

Debargha Roy [1], Pramit B. Chanda [1], Subir K. Sarkar [2]

1. *Computer Science and Engineering, Kalyani Government Engineering College, Kalyani, IND*
2. *Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, IND*

Received: May 25, 2026 | Review began: July 19, 2026 | Review ended: September 15, 2026 | Published: November 01, 2026

© **Copyright** 2026
This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Malaria is one of the most severe infectious diseases in the world, especially in tropical and subtropical areas, where timely diagnosis is essential for effective treatment. While microscopic examination of blood smear images is the gold standard, it is time-consuming, requires skilled personnel, and is prone to human error. Such restrictions have led to the emergence of automated computer-aided diagnostic systems based on deep learning frameworks with better performance.

This paper builds upon our earlier published paper on convolutional neural network (CNN)-based and Transformer-based malaria classification and proposes three hybrid deep learning architectures to detect malaria automatically. In addition to the standalone CNN and Transformer (Vision Transformer) models, the proposed model introduces (i) a CNN-Transformer-based hybrid model, (ii) a VGG19-Transformer hybrid model, and (iii) a ResNet-Transformer hybrid model. The goal of these architectures is to integrate the ability of convolutional networks to robustly extract local features with the ability of Transformer-based attention mechanisms to extract global features.

Each model is trained and tested on the available NIH Malaria Cell Image Dataset. Experimental results show that the proposed hybrid models outperform the standalone architectures, with the CNN-Transformer architecture achieving a maximum validation accuracy of around 95%. The results indicate that hybrid architectures offer a more comprehensive and efficient system for automated malaria diagnosis and show a promising direction for future medical image analysis research.

Categories: Computational Science and Engineering, Image Processing and Analysis, Medical Expert systems

Keywords: malaria, computer-aided diagnosis, blood smear images, deep learning, cnn, vision transformer, hybrid architectures, medical imaging, residual neural network(resnet), vgg-19

Introduction

Malaria remains a serious health issue in the world, especially in developing nations that are in tropical and subtropical regions. In 2022, the World Health Organization reported an estimated 249 million cases of malaria and nearly 608,000 deaths worldwide [1]. *Plasmodium* parasites are transmitted by infected female *Anopheles* mosquitoes. If left untreated, malaria may result in cerebral malaria, multiple organ failure, and death. Thus, prompt and precise detection of malaria parasites is still of the utmost significance in the context of good treatment and disease management.

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8

Currently, the diagnosis of malaria is primarily performed by examining images from blood smears. This approach is used in clinical practice, but it is time-consuming and involves manual observation by experienced pathologists. In addition, inconsistencies could occur due to differences in staining, image resolution, and interpretation by the human eye. Rapid diagnostic tests have also been implemented as another option, but in some instances, their accuracy is affected by environmental and biological conditions. These restrictions highlight the importance of effective and automatic diagnostic systems that can support healthcare practitioners.

The Transformer architecture, which was first introduced for natural language processing applications [2], has recently been extended to computer vision tasks, such as the Vision Transformer [3]. Unlike convolutional neural networks (CNNs), Transformers use self-attention mechanisms to identify relationships across the image. This ability makes Transformer-based models very useful for complex image analysis tasks. While these models have shown great potential in various medical imaging areas, their use in malaria detection has not yet been widely explored.

We previously tried to compare the standalone CNN model with the Transformer model for automatic malaria classification [4]. The CNN model achieved higher convergence speed and accuracy, while the Transformer model exhibited better global image relationship learning capability. However, both architectures had disadvantages in their standalone implementations.

Earlier automated malaria detection systems relied on traditional machine learning and image processing techniques, such as support vector machines, handcrafted feature extraction, and transfer learning-based approaches [5-7]. Deep learning has proven to be a powerful tool for medical image analysis in recent years. CNNs have shown state-of-the-art performance in image classification problems by learning spatial and textural information directly from image data. Good performance in detecting malaria parasites has been reported in several studies using CNNs [8-10]. However, CNN architectures are mostly concerned with local receptive fields and cannot model long-range contextual relationships in images. Some previous studies have also discussed challenges such as overfitting, limited dataset diversity, and generalization issues in automated malaria classification systems [11].

This extended work presents three hybrid deep learning architectures based on the interaction between convolutional and attention-based architectures to overcome these challenges. The proposed models are as follows:

• A CNN-Transformer hybrid for capturing local feature information and global attention mechanisms.

• The proposed ResNet-Transformer hybrid approach [12], featuring residual learning to enhance feature propagation and stability.

• A VGG19-Transformer hybrid model based on deep hierarchical feature representation and transfer learning [13,14].

This research aims to test whether hybrid architectures can be designed to achieve better performance while simultaneously capturing both the local and global characteristics of images. Moreover, recent research on hybrid image processing frameworks with Transformers has encouraged the development of such architectures [15]. Experimental results show that the proposed hybrid models yield superior performance compared with standalone CNN and Transformer architectures, demonstrating the effectiveness of hybrid transfer learning models for automated malaria diagnosis.

Materials And Methods

The main goal of the proposed methodology is to develop and test hybrid deep learning architectures for automated malaria parasite detection from thin blood smear images. In this work, we combine convolution-based feature extraction with Transformer-based attention mechanisms into a single hybrid framework and extend our previous comparative study [4] by implementing these architectures.

The methodology proposed in this work involves several steps: data acquisition, data preprocessing, hybrid model design, model training, and model evaluation. In this study, three hybrid architectures were developed: CNN-Transformer, VGG19-Transformer, and ResNet-Transformer. The overall workflow and architectural designs of the proposed hybrid

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8

models are shown in figures in the following sections.

Dataset description

The publicly available NIH Malaria Cell Image Dataset [16] was used for training and evaluation in this study. This dataset includes 27,558 segmented red blood cell images (13,738 images in each category).

Parasitized cells: Blood cells containing *Plasmodium* parasites.

Uninfected cells: Red blood cells that are not infected with malaria parasites.

The dataset is balanced, thus minimizing classification bias during training. The dataset was split into 70% for training, 20% for validation, and 10% for testing. To ensure a fair comparison among the models, all experiments were conducted using the same data partitioning.

Data preprocessing

Preprocessing is a crucial step in enhancing uniformity and the generalization performance of deep learning models. Before training, several preprocessing techniques were applied:

Image Resizing: To ensure a uniform input size across all architectures, all blood smear images were resized to 64×64 pixels.

Normalization: Pixel intensities were normalized to the range (0, 1) to help stabilize training and improve convergence.

Data Augmentation: Several augmentation techniques, such as random rotation, horizontal flipping, zoom transformation, and random shifting, were employed to minimize overfitting and make the models more robust. These augmentation techniques increased the variability of the training data and helped the models generalize across different image conditions.

Proposed hybrid architectures

CNN-Transformer Hybrid Architecture

The CNN-Transformer architecture was designed to fuse the feature extraction ability of CNNs with the global attention mechanism of Transformers. In this architecture, blood smear images are initially processed by convolutional layers to extract local features, including parasite texture, blood smear staining patterns, and cell morphology.

These feature maps are then converted into embeddings and fed into the Transformer encoder layers, which consist of Multi-Head Self-Attention layers and feed-forward neural network blocks. The self-attention mechanism allows the model to focus on relationships between different parts of the image over long distances.

Finally, dense layers followed by a sigmoid activation function are used for the binary classification of parasitized and uninfected cells. The CNN-Transformer hybrid architecture enhances feature representation and classification performance by combining local and global feature learning. The complete architecture of the proposed CNN-Transformer hybrid model is illustrated in Figure *1*.

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8

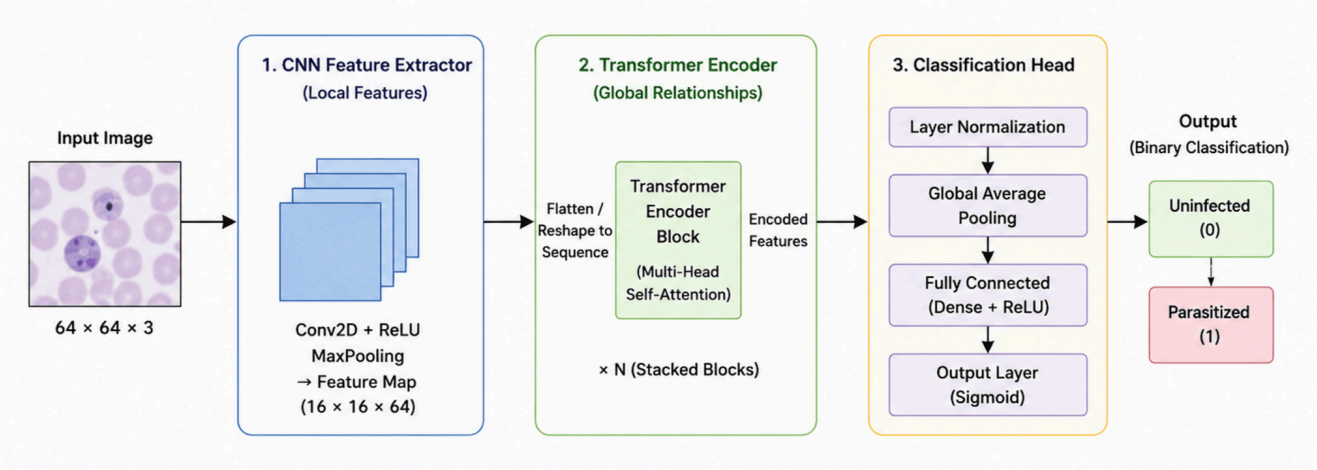


FIGURE 1: CNN-Transformer Hybrid Model for Malaria Detection

CNN, Convolutional Neural Network; ReLU, Rectified Linear Unit

ResNet-Transformer Hybrid Architecture

The ResNet-Transformer hybrid [12] combines residual learning with a Transformer-based attention mechanism to better represent features and provide learning stability.

Here, ResNet convolutional blocks were used to learn discriminative spatial features, and residual connections were used to achieve better gradient propagation and mitigate performance degradation in deeper networks. The obtained feature representations were then fed into the Transformer encoder layers to learn global context in the images.

The final classification layer was used for binary prediction (infected vs. uninfected) of blood cells. By combining residual learning with a self-attention mechanism, the architecture was able to achieve good performance and stable convergence during the training process. The overall design of the proposed ResNet-Transformer hybrid architecture is shown in Figure *2*.

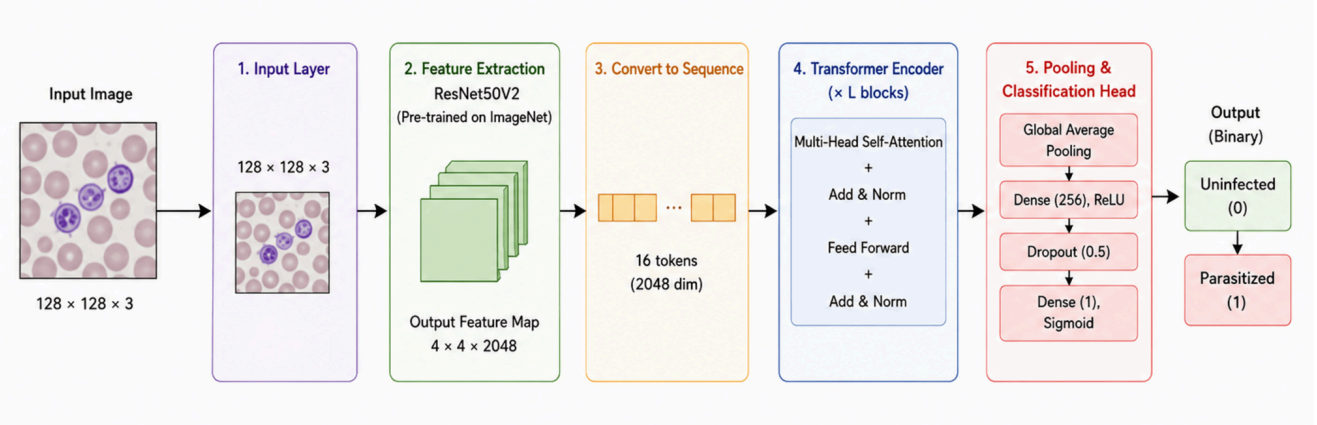


FIGURE 2: ResNet-Transformer Hybrid Model for Malaria Detection

ReLU, Rectified Linear Unit

VGG19-Transformer Hybrid Architecture

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8

This is a hybrid architecture combining a pre-trained VGG19 network [13] with Transformer encoder layers, called VGG19-Transformer. This model uses the VGG19 backbone as a deep hierarchical feature extractor that can capture detailed image representations of blood smear images.

To enhance the efficiency of feature extraction, transfer learning [14] was employed to take advantage of the weights pre-trained on VGG19, thereby reducing the complexity of training. The deep feature maps extracted from the network were then passed into Transformer encoder blocks to capture global contextual relationships in the images using the self-attention mechanism.

The final classification step involved a fully connected dense layer with a sigmoid activation function for binary output prediction. The model combines the contextual learning power of Transformers with the deep feature extraction power of VGG19. The detailed architecture of the VGG19-Transformer hybrid model is presented in Figure *3*.

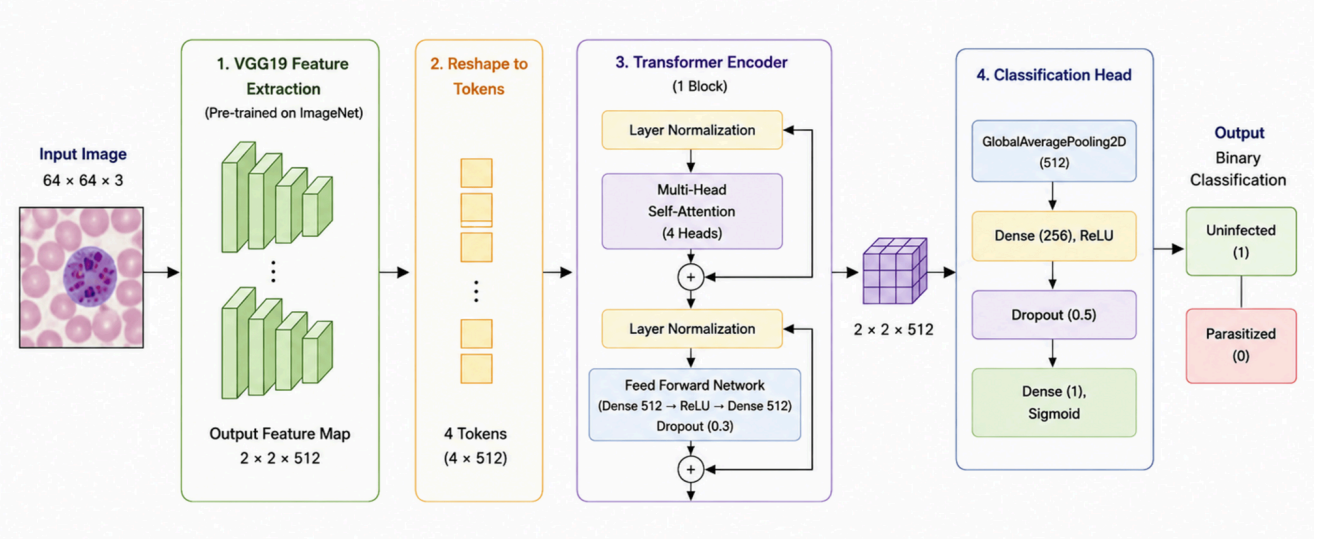


FIGURE 3: VGG19-Transformer Hybrid Model for Malaria Detection

Training strategy

To facilitate a fair comparison and reliable evaluation, all hybrid architectures were trained under the same experimental conditions.

The training settings were as follows:

- Optimizer: Adam
- Loss Function: Binary Cross-Entropy
- Batch Size: 32
- Epochs: 10
- Adaptive Learning Rate Reduction: Learning rate scheduling with adaptive learning rate reduction
- Early Stopping: Used to avoid overfitting

The same pipeline and training setup were used for all experiments. All the proposed hybrid architectures were implemented and trained using the TensorFlow deep learning framework [17].

Evaluation metrics

Several performance measures were used to evaluate the proposed models and provide a thorough analysis of their classification ability and robustness. The following metrics were used for evaluation: accuracy, precision, recall (sensitivity), F1-score, and confusion matrix. These metrics allowed the authors to compare the proposed hybrid architectures in terms of their ability to learn features, classify correctly, and generalize for the automatic detection of malaria parasites.

Artificial intelligence assistance

Large language model-based tools were used during the preparation of this manuscript for language refinement, grammar correction, paraphrasing, and readability improvement. The authors reviewed, verified, and edited all generated content and take full responsibility for the accuracy and integrity of the manuscript.

Results And Discussion

The training and testing of the proposed hybrid architectures were conducted on the NIH Malaria Cell Image Dataset [16] using the same experimental setup. To analyze performance, training accuracy, validation accuracy, training loss, and validation loss were calculated over 10 epochs. The main goal of the analysis was to investigate the performance of various hybrid architectures by incorporating convolutional blocks for feature extraction and Transformer blocks for global attention mechanisms in automated malaria detection.

Performance of the CNN-Transformer hybrid model

The CNN-Transformer hybrid model exhibited robust learning ability and convergence performance throughout the training process. This model incorporated convolutional layers to extract local features and Transformer encoder layers to capture global contextual information. The architecture achieved a final training accuracy of 96.13% and a final validation accuracy of 94.45%.

The validation accuracy remained between 94% and 95% for most of the training epochs, suggesting a good ability to generalize. The model also achieved relatively low validation loss, indicating that it did not overfit and thus learned effectively.

The CNN successfully represented local parasite features, such as staining texture and cell morphology, while the Transformer module captured contextual information by modeling long-range relationships within images.

The epoch-wise performance of the CNN-Transformer model is summarized in Table *1*, and the corresponding training and validation accuracy curves are shown in Figure *4*.

Hybrid CNN-Transformer Model (Epoch 1-10)				
Iteration	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.5699	0.6380	0.9495	0.1524
2	0.9567	0.1543	0.9515	0.1554
3	0.9593	0.1378	0.9485	0.1530
4	0.9594	0.1408	0.9492	0.1680
5	0.9578	0.1428	0.9475	0.1622

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8

6	0.9602	0.1329	0.9492	0.1672
7	0.9628	0.1286	0.9503	0.1648
8	0.9595	0.1313	0.9499	0.1594
9	0.9616	0.1254	0.9505	0.1592
10	0.9613	0.1232	0.9445	0.1853

TABLE 1: Iteration-Wise Performance of the Hybrid CNN-Transformer Model

CNN, Convolutional Neural Network

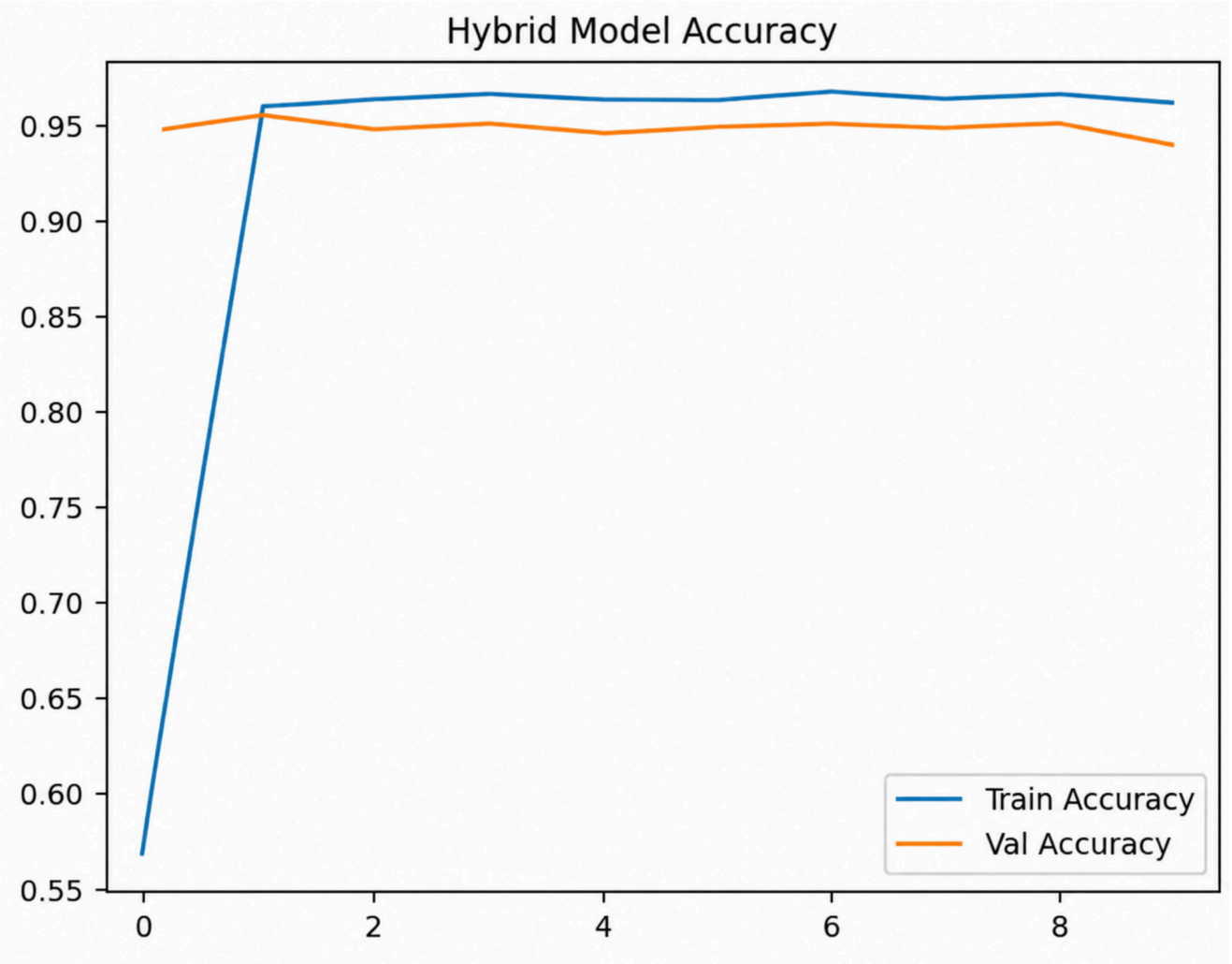


FIGURE 4: Accuracy Plot of the CNN-Transformer-Based Hybrid Framework

CNN, Convolutional Neural Network

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8

The accuracy graph of the CNN-Transformer hybrid model shows stable convergence behavior during the training process. The training and validation accuracies show steady improvement over the epochs, with relatively small fluctuations, suggesting effective feature learning and good generalization. The combination of convolutional feature extraction and Transformer-based attention mechanisms allowed the model to effectively learn both local parasite features and global contextual information from blood smear images.

Performance of the ResNet-Transformer hybrid model

The ResNet-Transformer hybrid architecture exhibited convergence and a gradual improvement in performance over training and validation epochs. The architecture achieved a training accuracy of 94.19% and a validation accuracy of 93.03%.

The introduction of residual connections helped improve the gradient and allowed stable training of deeper convolutional layers. The accuracy of the validation set increased steadily during the training process, whereas the validation loss continuously decreased from 0.2971 to 0.1815, which showed that the optimization behavior was stable.

The learning ability was strong in ResNet-Transformer, and the feature propagation was enhanced by the residual learning, though the validation accuracy was slightly lower than that of the CNN-Transformer model and the VGG19-Transformer model.

The epoch-wise training and validation results of the ResNet-Transformer model are presented in Table *2*, and the training and validation accuracy trends of the model are illustrated in Figure *5*.

Hybrid ResNet-Transformer Model (Epoch 1-10)				
Iteration	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.7449	0.5699	0.8858	0.2971
2	0.8915	0.2997	0.9094	0.2452
3	0.9026	0.2671	0.9131	0.2382
4	0.9143	0.2306	0.9211	0.2148
5	0.9229	0.2147	0.9234	0.2127
6	0.9273	0.2005	0.9267	0.2006
7	0.9301	0.1894	0.9269	0.1960
8	0.9349	0.1732	0.9283	0.1887
9	0.9396	0.1703	0.9298	0.1830
10	0.9419	0.1627	0.9303	0.1815

TABLE 2: Iteration-Wise Output of Hybrid ResNet-Transformer Model

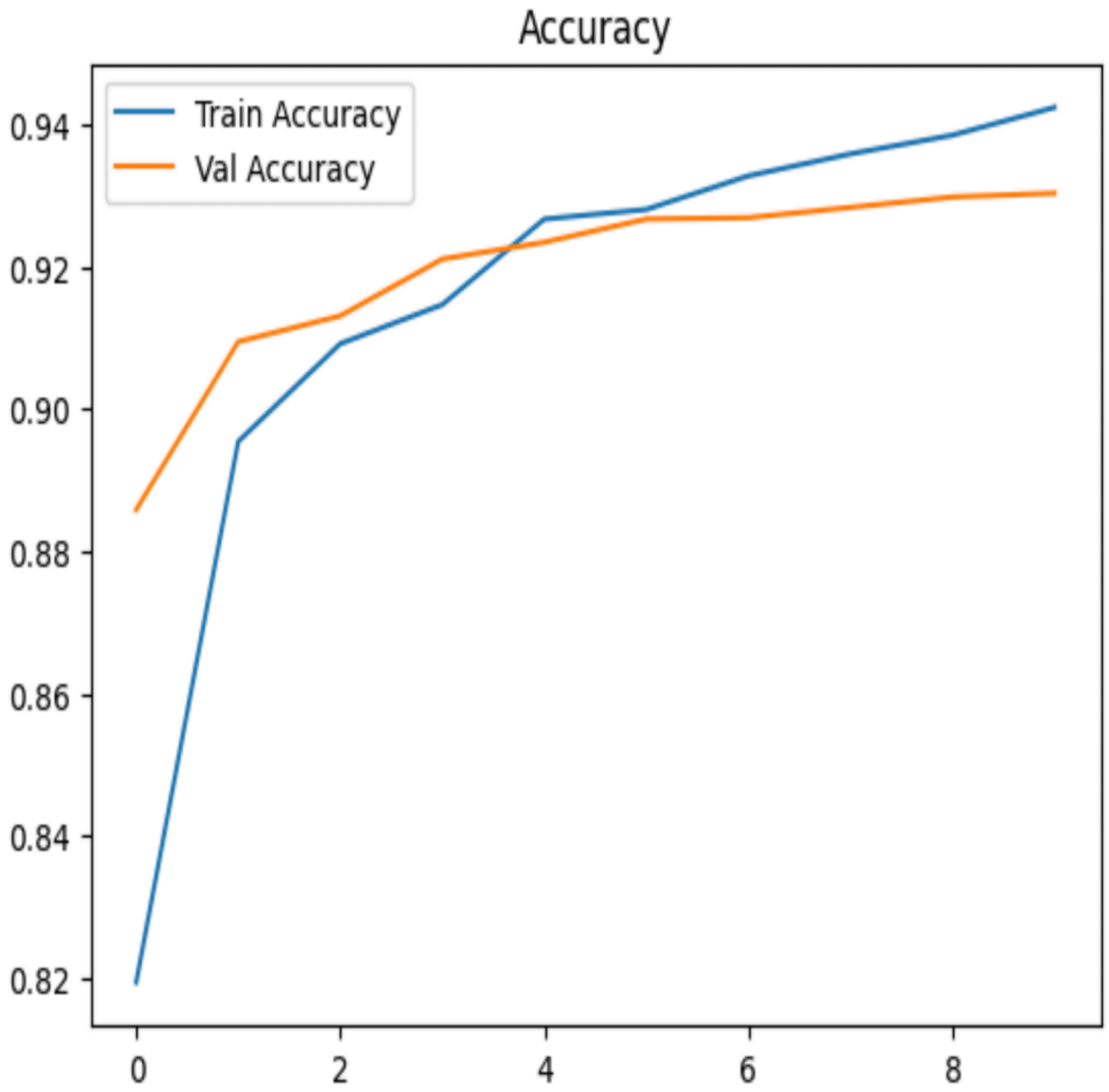


FIGURE 5: Accuracy Plot of the ResNet-Transformer-Based Hybrid Model

The accuracy graph of the ResNet-Transformer hybrid model showed a gradual and stable increase in training and validation accuracy. Residual connections helped smooth the optimization process and improve gradient flow through the deeper layers. The architecture also demonstrated good feature learning ability during experimentation, with performance remaining stable throughout training, although its final accuracy was slightly lower than that of the other hybrid models.

Performance of the VGG19-Transformer hybrid model

The overall classification performance of the VGG19-Transformer hybrid model was the highest among all the implemented architectures. By combining the transfer learning capabilities of VGG19 with self-attention mechanisms, the model achieved remarkable results in feature representation and contextual learning. The architecture achieved a final

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8

training accuracy of 97.03% and a final validation accuracy of 95.90%.

The model maintained consistently high validation accuracy throughout the training process, and its validation loss was the lowest among all hybrid architectures. The VGG19 network's deep hierarchical feature extraction capability significantly enhanced representation learning, whereas the Transformer encoder's global dependency modeling capability was supported by self-attention mechanisms.

The excellent results of this architecture suggest that transfer learning and Transformer-based contextual learning can significantly boost performance in automated malaria classification. The detailed epoch-wise performance metrics of the VGG19-Transformer model are summarized in Table *3*, and the corresponding accuracy curve is shown in Figure *6*.

Hybrid VGG19-Transformer Model (Epoch 1-10)				
Iteration	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.8767	0.2851	0.9485	0.1446
2	0.9534	0.1414	0.9475	0.1402
3	0.9572	0.1254	0.9544	0.1216
4	0.9617	0.1133	0.9550	0.1226
5	0.9632	0.1047	0.9534	0.1257
6	0.9659	0.1020	0.9490	0.1309
7	0.9678	0.0884	0.9548	0.1297
8	0.9722	0.0821	0.9574	0.1186
9	0.9715	0.0793	0.9601	0.1145
10	0.9703	0.0830	0.9590	0.1125

TABLE 3: Iteration-Wise Performance of the Hybrid VGG19-Transformer Model

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8

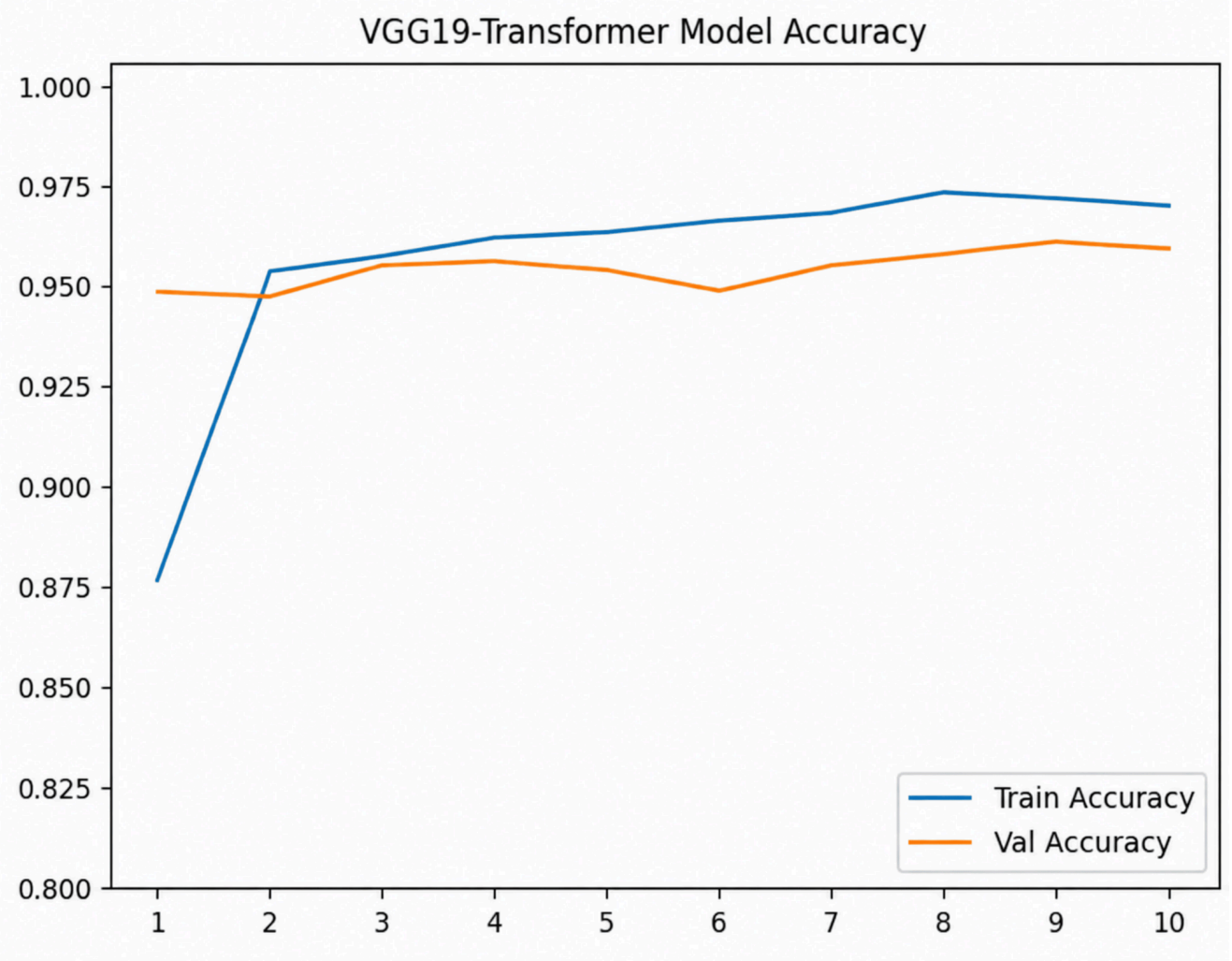


FIGURE 6: Accuracy Plot of VGG19-Transformer-Based Hybrid Model

The accuracy curves show rapid convergence, with high validation accuracy throughout the epochs and fewer fluctuations in validation loss. The deep hierarchical feature extraction provided by VGG19 and the global contextual modeling provided by Transformer self-attention together offered better classification ability and generalization performance for automated malaria detection.

Comparative performance analysis

The effectiveness of the proposed hybrid architectures was evaluated through simulations and compared with previously reported architectures for malaria detection, as well as with the standalone CNN and Vision Transformer models from our previous study [4]. Multiple parameters, such as validation accuracy, convergence behavior, computational complexity, and overall feature learning capability, were used for comparison.

The results of the experiments suggest that hybrid deep learning models achieve better classification accuracy by leveraging the strengths of both convolutional networks for local feature extraction and Transformers for global contextual understanding. A comparison of previously reported approaches, standalone architectures, and the proposed hybrid models is presented in Table *4*.

Serial No.	Model/Approach	Method Used	Validation Accuracy

1	Rajaraman et al. [18]	Pre-trained CNN	≈ 91%
2	Maqsood et al. [8]	Custom CNN	≈ 92%
3	Bibin et al. [19]	Deep Belief Network	≈ 90%
4	Previous Work [4]	Custom CNN	≈ 94%
5	Previous Work [4]	Vision Transformer	≈ 91.5%
6	**Proposed Work**	**CNN–Transformer Hybrid**	**≈ 94.5%**
7	**Proposed Work**	**ResNet–Transformer Hybrid**	**≈ 93%**
8	**Proposed Work**	**VGG19–Transformer Hybrid**	**≈ 96%**

TABLE 4: Comparative Performance Analysis of Existing and Proposed Models for Automated Malaria Detection

CNN, Convolutional Neural Network

From conventional deep learning methods to modern hybrid methods, the comparative analysis reveals a gradual increase in the performance of automated malaria classification. Competitive performance was previously reported for CNN-based techniques thanks to their ability to extract spatial features, while Transformer-based models excelled in contextual understanding through their self-attention mechanism.

In our previous work [4], we developed a standalone CNN with good local feature learning and efficient convergence behavior. The Vision Transformer, on the other hand, offered better overall contextual modeling capability, albeit at a higher cost in computational resources and training time.

The proposed hybrid architectures have demonstrated the ability to integrate the benefits of both convolutional and attention-based learning mechanisms. The hybrid models combined deep learning feature extraction and Transformer-based contextual learning, resulting in enhanced feature representation and generalization ability.

The VGG19-Transformer model appeared to be the most successful of all the architectures, with a validation accuracy of about 96% compared to the CNN and the Vision Transformer. This architecture showed good performance due to the synergy of transfer learning, deep hierarchical feature extraction, and self-attention-based contextual analysis.

The results of the experiments indicate that the hybrid deep learning architectures were more balanced and effective than standalone CNN or Transformer-based architectures for the automatic diagnosis of malaria.

Observational analysis

The Performance of Hybrid Architectures

The results of the experiments show that the hybrid deep learning architecture outperforms both the convolutional and Transformer models. The proposed models demonstrated enhanced robustness and classification ability by combining local feature extraction with global contextual learning.

CNN-Transformer Observations

The CNN-Transformer architecture showed balanced performance and stable convergence behavior. The architecture successfully represented the morphology and spatial texture of the parasites and included the modeling of larger image relationships.

VGG19-Transformer Observations

The VGG19-Transformer model achieved the highest overall accuracy among all the architectures. The Transformer encoder was used to introduce contextual information into blood smear images, while transfer learning was used to extract deep features efficiently. This architecture performs much better than the other models in terms of generalization and training loss.

ResNet-Transformer Observations

The residual learning mechanism of the ResNet-Transformer architecture had a positive effect on training stability and gradient propagation. Its classification accuracy was slightly lower than that of the other hybrid models, but the architecture exhibited good robustness and consistent optimization.

Discussion

The experimental results show that combining CNNs with Transformer-based attention mechanisms can significantly boost the performance of automated malaria detection. The proposed hybrid architectures were able to integrate local spatial feature extraction, deep hierarchical representation learning, and global contextual dependency modeling.

The VGG19-Transformer hybrid model resulted in the highest overall performance among all proposed models, combining the advantages of transfer learning and self-attention-based contextual analysis. Based on these results, it is clear that hybrid deep learning architectures are a very promising path for future medical image analysis systems. Learning local information and global contextual relationships at the same time can greatly enhance the accuracy of automated classification of malaria.

Special Mention: In addition to the major field of automated classification of malaria, recent interdisciplinary research has illustrated the use of high-level analytical, computational, structural, sensing, and environmental treatment methods in a variety of scientific disciplines, such as supramolecular assemblies, coordination frameworks, computational structural analysis, molecular sensing, and wastewater treatment systems. The studies presented here are but examples of the use of specialized scientific methods to solve research problems that are more challenging, and show how computational, analytical, and experimental techniques can be brought together in an interdisciplinary fashion across scientific fields [20-25].

Conclusions

The study presented in this research is an integrated investigation of hybrid deep learning architectures for automated malaria classification based on thin blood smear images. The work extends our previous comparative study of CNN-based and Transformer-based models by designing hybrid models that perform feature extraction using CNNs and contextual analysis using Transformer-based attention mechanisms for improved medical image classification. The experimental results showed that the proposed hybrid architectures achieved better performance by learning both local spatial features and global contextual information from blood smear images, consequently improving classification performance. Transfer learning, residual learning, and self-attention mechanisms were integrated to achieve better feature representation, more stable convergence behavior, and improved generalization ability during training. The model combining deep transfer learning and Transformer-based contextual modeling achieved the highest classification accuracy (around 95.90% on the validation set) among all implemented architectures. The experimental results further showed that hybrid models are more robust than single-convolutional or attention-based models and exhibit less overfitting.

The study also demonstrated that the proposed CNN-Transformer architecture is more effective for automated malaria diagnosis than either CNN or Transformer alone. The proposed method was able to capture detailed information about parasite morphology and the broader context of blood smear images. This study demonstrates that hybrid deep learning has great potential for developing reliable computer-based diagnostic models for malaria classification. These architectures could play a role in developing faster, more accurate, and automated medical diagnosis systems, especially

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8

in resource-limited healthcare settings where early diagnosis of malaria remains a critical issue. The proposed architectures demonstrated good classification results, but some improvements and extensions can be considered in the future to further improve diagnostic performance and applicability. Future research could focus on extending the system for multiclass malaria classification involving different parasite life-cycle stages, using larger and more diverse datasets acquired from different clinical sources, incorporating explainable artificial intelligence approaches to increase interpretability and clinical trust, and optimizing the architectures for deployment on mobile and edge devices for real-time malaria screening. As hybrid deep learning systems continue to evolve, they can substantially enhance computer-aided malaria diagnosis and contribute to the development of intelligent, efficient, and accessible healthcare solutions.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Debargha Roy, Pramit B. Chanda, Subir K. Sarkar

Acquisition, analysis, or interpretation of data: Debargha Roy

Drafting of the manuscript: Debargha Roy

Critical review of the manuscript for important intellectual content: Debargha Roy, Pramit B. Chanda, Subir K. Sarkar

Supervision: Pramit B. Chanda, Subir K. Sarkar

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue.

Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue.

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following:

- **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work.
- **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work.
- **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

Acknowledgements

I want to extend my sincere thanks to all those who contributed towards making this research possible. I also owe special thanks to the creators of the NIH Malaria Cell Image Dataset [16] and Kaggle for providing access to this dataset, which was crucial to ensuring reliability and consistency when training and testing the proposed hybrid deep learning models. Gratitude is also due to the developers and contributors of TensorFlow and Keras [17], whose frameworks and tools were the basis for implementing and experimenting with the proposed architectures during this research work. I wish to extend my special thanks to Prof. Pramit Brata Chanda Sir, my mentor, who always guided and encouraged me and gave valuable suggestions that helped greatly in the successful completion and overall direction of this research. The deep learning, model evaluation, and comparative analysis aspects of this work were greatly aided by his inputs. Special thanks

are extended to the ICRCICN 2025 Conference, at which we submitted our research paper. The recognition and exposure gained at the conference inspired and motivated the continuation of this extended research work in the field of automated medical image analysis. In addition, I would like to express my sincere gratitude to the research and academic community and to the faculty members of the Department of Computer Science and Engineering, Kalyani Government Engineering College (KGEC), for their motivational and inspirational support, encouragement, and guidance in carrying out this research work in the field of medical image analysis and deep learning.

References

1. World Health Organization: World Malaria Report 2022. World Health Organization, Geneva, Switzerland; 2022.
2. Vaswani A, Shazeer N, Parmar N, et al.: Attention is all you need. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017, 6000-6010.
3. Dosovitskiy A, Beyer L, Kolesnikov A, et al.: An image is worth 16x16 words: Transformers for image recognition at scale [PREPRINT]. arXiv. 2021, 10.48550/arXiv.2010.11929
4. Roy D, Chanda PB, Sarkar SK: Comparative analysis on automated malaria classification using image processing and deep learning methodology. 2025 Seventh International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kalyani, India. 2025, 78-83. 10.1109/ICRCICN68210.2025.11364861
5. Das DK, Ghosh M, Pal M, Maiti AK, Chakraborty C: Machine learning approach for automated screening of malaria parasite using light microscopic images. Micron. 2013, 45:97-106. 10.1016/j.micron.2012.11.002
6. Díaz G, González FA, Romero E: A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. Journal of Biomedical Informatics. 2009, 42:296-307. 10.1016/j.jbi.2008.11.005
7. Fuhad KMF, Tuba JF, Sarker MRA, Momen S, Mohammed N, Rahman T: Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application. Diagnostics. 2020, 10:329. 10.3390/diagnostics10050329
8. Maqsood A, Farid MS, Khan MH, Grzegorzek M: Deep malaria parasite detection in thin blood smear microscopic images. Applied Sciences. 2021, 11:2284. 10.3390/app11052284
9. Poostchi M, Silamut K, Maude RJ, Jaeger S, Thoma G: Image analysis and machine learning for detecting malaria. Translational Research. 2018, 194:36-55. 10.1016/j.trsl.2017.12.004
10. Rosado L, Correia da Costa JM, Elias D, Cardoso JS: Automated detection of malaria parasites on thick blood smears via mobile devices. Procedia Computer Science. 2016, 90:138-144. 10.1016/j.procs.2016.07.024
11. Shambhu S, Koundal D, Das P, Hoang VT, Tran-Trung K, Turabieh H: Computational methods for automated analysis of malaria parasite using blood smear images: Recent advances. Computational Intelligence and Neuroscience. 2022, 2022:1-18. 10.1155/2022/3626726
12. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016, 770-778. 10.1109/CVPR.2016.90
13. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition [PREPRINT]. arXiv. 2015, 10.48550/arXiv.1409.1556
14. Pan SJ, Yang Q: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering. 2010, 22:1345-1359. 10.1109/tkde.2009.191
15. Chen H, Wang Y, Guo T, et al.: Pre-trained image processing transformer. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA. 2021, 12294-12305. 10.1109/CVPR46437.2021.01212
16. Malaria Cell Image Dataset. (2018). Accessed: October 27, 2026: https://ceb.nlm.nih.gov/repositories/malaria-datasets/.
17. TensorFlow: An end-to-end platform for machine learning. Accessed: October 27, 2026: https://www.tensorflow.org/.
18. Rajaraman S, Jaeger S, Antani SK: Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. PeerJ. 2019, 7:e6977. 10.7717/peerj.6977
19. Bibin D, Nair MS, Punitha P: Malaria parasite detection from peripheral blood smear images using deep belief networks. IEEE Access. 2017, 5:9099-9108. 10.1109/access.2017.2705642

20. Shankar R, Mendiratta S, Jakhar E, Singh R, Kaur Jassal A, Kociok-Köhn G: A rational synthesis of ladder-like motif in zinc-methylphosphonate from a preformed coordination assembly. Inorganica Chimica Acta. 2018, 482:681-686. 10.1016/j.ica.2018.07.016

21. Shankar R, Singh R, Kociok-Köhn G, Molloy KC: Studies on the coordination behavior of methanesulfonate in zinc(II) based two and three dimensional supramolecular assemblies. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences. 2014, 84:157-163. 10.1007/s40010-013-0121-5

22. Shankar R, Singh R, Kociok-Köhn G, Frontera A: A combined experimental and computational study of a supramolecular assembly based on cationic zinc(II)-ethanesulfonate. Journal of Molecular Structure. 2020, 1202:127206. 10.1016/j.molstruc.2019.127206

23. Singh R, Kociok-Köhn G, Kaur Jassal A, Singh L: Three-dimensional hydrogen-bonded magnesium(II) supramolecular motifs based on in situ generated alkanesulfonate (Me/Et/nPrSO3–) ligands: A combined experimental and computational study. Polyhedron. 2020, 175:114200. 10.1016/j.poly.2019.114200

24. Muwal PK, Saini AK, Choudhary K, et al.: Development of cyclic and acyclic fluorescent chemosensors for selective detection of ions and biomolecules in water. ChemistrySelect. 2025, 10:e202404394. 10.1002/slct.202404394

25. Agarwal M, Kumar R, Saxena LK, Tripathi A, Singh R: Comparative analysis of the treatment potential of cationic starch and its polyacrylamide blend on tannery and sewage effluents. Asian Journal of Chemistry. 2025, 37:2532-2544. 10.14233/ajchem.2025.34437

How to cite this article:
Roy D, Chanda P B, Sarkar S K (November 01, 2026) Implementation of a Hybrid Deep Learning Framework for Automated Malaria Classification: Integrating Convolutional Neural Network, ResNet, and Transformer Models . Cureus J Comput Sci 3 : es44389-026-00349-8. DOI https://doi.org/10.7759/s44389-026-00349-8