Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 3, 2026

ENHANCING MALARIA PARASITE CLASSIFICATION USING THIN BLOOD SMEAR IMAGES THROUGH DEEP LEARNING OPTIMIZER SELECTION

Muhammad Talha Jahangir[*1], Samina Naz[2], Zaid Sarfraz[3], Muhammad Rehman[4], Anas Mahmood[5]

[*1,3,4,5]*Dept. of Computer Science, Muhammad Nawaz Sharif University of Engineering & Technology, Multan, Pakistan*
[2]*Faculty of Computing and Emerging Technologies, Emerson University Bosan road Multan*

[*1]mtalhajahangir@mnsuet.edu.pk

DOI: http://doi.org/10.5281/zenodo.21484706

Keywords:
Malaria Detection, Thin Blood Smear Images, EfficientNetB4, Deep Learning, Transfer Learning, Medical Image Classification, Optimization Algorithms, Adam, SGD, RMSprop, Computer-Aided Diagnosis.

Article History
Received: 31 January 2026
Accepted: 16 March 2026
Published: 30 March 2026

Copyright @Author
Corresponding Author: *
Muhammad Talha Jahangir

Abstract

Malaria is still one of the most widespread vectors borne diseases in the world. Generally, the diagnosis is done manually by viewing thin blood smear images on a microscope, which is time-consuming, labor-intensive and human error prone. However, with recent developments in deep learning, automatic diagnosis systems can be developed to improve the accuracy and speed of diagnosis. In this research, an automatic Malaria Cell Detection Framework using EfficientNetB4 Convolutional Neural Network (CNN) Architecture is presented which is compared with the different optimization methods to enhance the classification accuracy. Pre-processing images, data augmentation, transfer learning and fine-tuning the EfficientNetB3 model are suggested to be used. A variety of optimization algorithms such as Adam, SGD with momentum and RMSprop are tested using important metrics like accuracy, precision, and sensitivity. Recall, F1 score and Area Under Receiver Operating Characteristic Curve (AUC-ROC). Experimental results show that the model convergence and stability as well as generalization performance can vary greatly depending on the type of the optimizer used. The Adamax optimizer outperforms the other optimization methods in terms of classification accuracy with 97.57%. The results reveal the efficacy of EfficientNetB4 model as a light-weight yet efficient network for malaria parasite detection in thin blood smear images and informative guidelines for the selection of suitable optimization techniques for medical image classification tasks. The findings of this study can be used for the development of robust computer aided diagnostic systems which provide early detection of malaria in a health care setting, especially in resource-limited environments.

1. INTRODUCTION

Malaria is still one of the most important vector-borne diseases and is a public health problem of great importance, especially in the tropics and subtropics. According to the World Health Organization (WHO) World Malaria Report 2025 [1], in 2024 there were an estimated 282 million malaria cases reported in the 80 malaria-endemic countries of the world, and an estimated 610,000 deaths from malaria. Almost 95% of the global malaria burden was in the WHO African

Region, with around 95% of all malaria cases and deaths occurring in children under 5 years of age and who made up around 75% of all malaria-related deaths in the region. Although malaria has been significantly reduced in terms of disease burden since 2000 with an estimated 2.3 billion cases and 14 million deaths prevented, the disease still has a substantial disease burden on health systems in many LMICs and there is a critical need for accurate, fast and cost-effective diagnostic disease detection technologies.

Malaria continues to be one of the highest burden countries in the WHO Eastern Mediterranean Region for Pakistan. Malaria re-emerged in the country after the devastating floods caused by climate change in 2022, as confirmed cases rose from nearly 399,097 in 2021 to nearly 2.7 million in 2023 [2]. In 2025, Pakistan estimated 10% reduction in malaria incidence when compared with 2024; but with a reported malaria incidence of 1.8 million cases, malaria is still a serious public health problem especially in rural and un-served areas where access to expert microscopy is limited. These challenges highlight the need to create computer-aided diagnostic (CAD) automated systems that can facilitate timely and accurate and even scalable malaria diagnosis.

Deep learning models trained on labelled microscopy datasets have been found to be as or more accurate than expert microscopists in several controlled settings across a variety of pathology domains. When labelled medical imaging data is sparse, transfer learning, where feature representations extracted from large scale datasets like ImageNet are tuned to specific domain classification tasks, has been effective in particular [3]. Convolutional neural networks (CNNs) have created a new branch of the automated analysis of medical images. EfficientNetB4 is a member of the EfficientNet series proposed by Tan and Le [4] that provides SOTA performance with compound scaling of network depth, width and input resolution. Although there is a lot of literature on the CNN for malaria detection, there are few systematic comparisons of optimisation algorithms on EfficientNetB4. The type of optimiser that is used has a great impact on convergence speed, generalisation power and the potential for overfitting. Other algorithms, like AdamW, Adamax, RMSprop, and SGD have been tried and tested as alternatives to Adam, and have unique dynamical properties that can be beneficial in certain loss landscapes [5].

The paper introduces an automatic malaria parasite classification system using AI and deep learning techniques on thin blood smear images. The proposed solution is based on the EfficientNetB4 architecture, transfer learning and investigates the impact of optimization methods on classification. An extensive comparison among Adam, AdamW, Adamax, RMSprop and SGD is conducted with identical experimental setup to determine which one works best. From the experimental results, it is observed that the optimizer named Adamax is performing well in all the experiments and it is the best optimizer for malaria parasite classification with reliable and accurate results. To sum up, the following contributions to the field were made in this study:

1. To accurately classify the thin blood smear images, the EfficientNetB4 transfer learning model is proposed for development of an AI-powered malaria parasite classification framework.
2. Five popular optimization algorithms (SGD, RMSprop, Adam, AdamW, and Adamax) are compared in detail to see how they affect model convergence and classification accuracy.
3. Proposed framework has a competitive accuracy of 97.57% which is better than existing methods reported in literature.

This remainder of this paper is structured as follows. A brief summary of recent research in automatic detection of malaria parasite and deep learning based classification of medical images is summarized in Section 2. In this Section, the dataset, preprocessing, data augmentation, EfficientNetB4 based architecture and optimizer configurations are discussed followed by the evaluation metrics. The methodology used in the experimental setup and a detailed performance comparison of the optimization algorithm

evaluated is described in section 4. The results of the experiments are discussed in Section 5, and the constraints of the proposed framework are pointed out. Finally, Section 6 concludes the paper and offers some directions for future research.

2. RELATED WORK

Malaria is still one of the major public-health challenges in the world with hundreds of millions of cases reported annually. Manual examination of thin and thick blood smears stained with the dye Giemsa is the clinical reference standard for the diagnosis with a trained microscopist visually looking for Plasmodium parasites in red blood cells. This procedure is reliable, but time consuming and labor-intensive, and may be subject to inter- and intra-observer variability, particularly in resource-poor areas where experienced microscopists are limited. The constraints have encouraged automatic computer-aided diagnosis (CAD). This has become the "gold standard" for the task, with the public NIH/NLM malaria dataset consisting of 27,558 images of segmented red blood cell sections from thin blood smears (13,779 with and 13,779 without parasites). The latest research (2022-2026) has focused on custom and lightweight convolutional neural networks (CNNs), and even on aggressive data augmentation. Optimised by reinforcement learning, a data-augmentation CNN (DACNN) achieved high accuracy of 94.79% on the NIH balanced dataset [2]. Magotra and Rohil [4] created a light CNN that achieved an accuracy of approximately 96.0% and was fast to train compared to VGG-19 and Inception-v3. Hoyos and Hoyos [3] used a YOLOv8 detector with data augmentation to both locate and count parasites, achieving a detection accuracy of 95.0% and significantly cutting down the time spent on analysis. The current paradigm is transfer learning from image-net pre-training. Ozsahin et al. [5] tested a CNN on thin and thick peripheral smears, and obtained 96.97% accuracy in the case of the thick smears. Hemachandran et al. [6] took a custom CNN and compared it with ResNet-50 and MobileNetV2, with the best accuracy being obtained using MobileNetV2 (97.06%). The DenseNet-201 outperformed other off-the-shelf architectures and remained the most stable and robust with a 97.45% result in the cross-dataset experiments [7]. Nusrat et al. [8] compared six deep networks and concluded that XceptionNet is the best one with 97.55%. On the other hand, Laghari et al. [1] reported that fine-tuning of a handful of pretrained CNNs (ResNet-50, ResNet-101, Xception) resulted in a best accuracy of 89% showing the importance of proper fine-tuning and selection of the optimiser. In a recent study, Jones and Murugamani (2023) employed a pipeline of denoising the blood smear images using the bilateral filter, extracting features from the images using a CNN, reducing the dimensions of such features using an updated grey-wolf optimization algorithm, and finally, a shallow classifier, SVM, for classification; they tested their pipeline on the NIH malaria dataset and reported an accuracy of around 94% with competitive precision, recall, F-measure and specificity, suggesting that an end-to-end CNN with a shallow-classifier might be competitive with a deeper end-to-end network while being more efficient at inference. A related comparative study of 13,750 images of cell images containing and not containing parasites that were generated from a set of cells developed by the NIH. Shambhu et al., (2022) perform a detailed review on the entire computational process of malaria diagnosis from literature, including image acquisition, image preprocessing, segmentation of the red blood cells, feature extraction and selection, and classification methods. This is not a report of original experimental results, but a synthesis of recurring design decisions and open challenges; this review is used here mainly to support the trends that emerge from the experimental studies above, especially the persistence of the NIH dataset as the primary source of experimental data; and the continued lack of experimental results that are ready for clinical deployment. Impressive as this accuracy (99.98%) is, Umer et al. (2024) achieve it with a novel Stacked CNN architecture on the 27,558-image NIH set under 5-fold cv, results which should be examined for possible overfitting to the

relatively homogeneous NIH dataset and for lack of external validation. [12] Mujahid et al. (2024) applied EfficientNet with transfer learning and k-fold cross-validation on the NIH/Kaggle malaria cell-image dataset, achieving 97.57% accuracy and outperforming several other pre-trained CNN backbones, which further underscores the value of transfer learning in this domain. Beyond binary classification, several recent works target richer outputs such as cell localization or species-level identification. The present work fine-tunes an ImageNet-pretrained EfficientNet-B4 network—augmented with batch normalization and dropout regularization—and trains it with the Adamax optimiser (learning rate 0.001). On the NIH thin-blood-smear dataset, the proposed model achieves 97.57% accuracy, 97.75% sensitivity, 97.39% specificity, 97.40% precision, an F1-score of 97.57%, and an AUC of 0.9962, thereby outperforming the recent machine-learning and deep-learning methods summarized in Table 1. The closely balanced sensitivity and specificity indicate reliable detection of both parasitized and uninfected cells—a desirable property for an automated malaria-screening tool.

Table 1. Comparison with recent (2022–2026) malaria-detection methods reporting accuracy below the proposed model.

Study [Ref.]	Year	Approach / Model	Dataset	Accuracy (%)
Laghari et al. [1]	2025	Transfer learning (ResNet-50 / ResNet-101)	NIH malaria	89.00
Oyewola et al. [2]	2022	Data-augmentation CNN (DACNN)	NIH malaria (Kaggle)	94.79
Hoyos & Hoyos [3]	2024	YOLOv8 detector + augmentation	Own blood smears	95.00
Magotra & Rohil [4]	2022	Lightweight custom CNN	NIH malaria	96.00
Ozsahin et al. [5]	2022	Custom CNN (thick smear)	NIH malaria (thick)	96.97
Hemachandran et al. [6]	2023	MobileNetV2 (transfer learning)	NIH malaria	97.06
Loddo et al. [7]	2022	DenseNet-201 (transfer learning)	NIH malaria	97.45
Nusrat et al. [8]	2025	XceptionNet (transfer learning)	NIH malaria	97.55
Jones and Murugamani . [9]	2023	Bilateral filtering + CNN features + SVM	NIH malaria	94.00
Shambhu et al. [10]	2022	YOLOv4, Faster R-CNN, SSD300	NIH malaria	93.87
Umer et al. . [11]	2024	Stacked CNN, 5-fold cross-validation	NIH malaria	96.98
Mujahid et al.	2025	Transfer learning (EfficientNet)	NIH/Kaggle malaria dataset	97.33

3. PROPOSED METHODOLOGY

In this section, the entire methodology of the proposed malaria cell-detection framework is described. The entire pipeline is broken down into five stages: (i) acquisition and stratified partitioning of the NIH Malaria Cell Image Dataset; (ii) image pre-processing, conservative training-stage augmentation; (iii) an EfficientNetB4 backbone pre-trained on ImageNet and fine-tuned end-to-end, followed by an integrated classification head; (iv) a controlled comparison of five optimisation algorithms under identical experimental conditions; and (v)

evaluation using six diagnostic metrics, followed by deployment as an interactive Streamlit web application of the best-performing model.

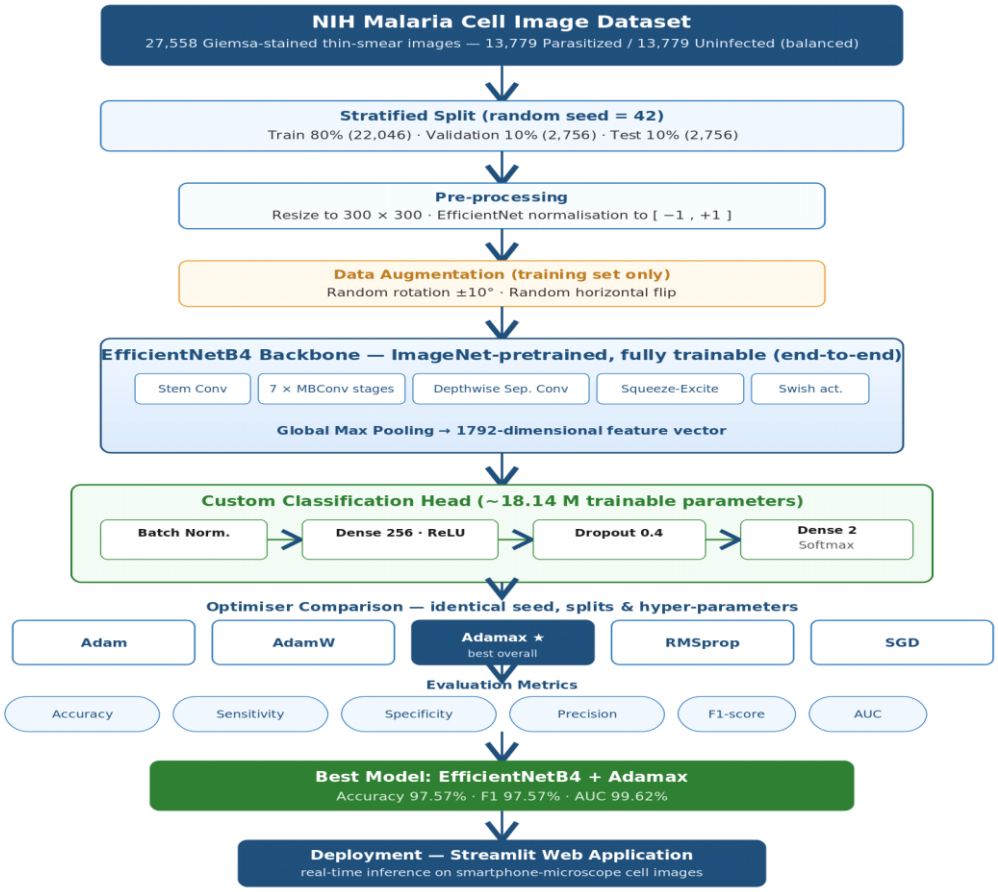


Figure 3. This presentation will provide a step-by-step overview of the proposed EfficientNetB4 cell-detection methodology starting from the NIH dataset and stratified split, moving on to pre-processing and augmentation, the completely fine-tuned EfficientNetB4 backbone with a custom classification head, the controlled 5-optimizers comparison and evaluation, and finally the deployment of the best (Adamax) model as a Streamlit web application.

3.1 Dataset

All experiments were performed on the publicly available NIH Malaria Cell Image Dataset, hosted by the National Institutes of Health (NIH) and the National Institute of Allergy and Infectious Diseases (NIAID) [40]. The dataset contains 27,558 segmented images of individual red blood cells extracted from Giemsa-stained thin blood-smear slides, evenly divided into two classes: Parasitized (13,779 images) and Uninfected (13,779 images with healthy erythrocytes). Because the dataset is perfectly balanced, no class-weighting or over-sampling techniques were required. The data were partitioned into training (80%), validation (10%), and test (10%) subsets using stratified splitting with a fixed random seed (42) to guarantee reproducibility, yielding 22,046 training, 2,756 validation, and 2,756 test images.

3.2 Data Pre-processing and Augmentation

All images were resized to 300 × 300 pixels, the native input resolution of EfficientNetB4, and pixel intensities were normalised with the EfficientNet-specific pre-processing function, scaling values to the [−1, +1] range in accordance with the ImageNet standardisation protocol.

Following established practice for limited medical-imaging datasets, mild augmentation was applied to the training set only, comprising random rotations in the range −10° to +10° and random horizontal flips. These transformations are conservative in nature and can be used to add diversity into the training set, but without impacting the morphology of the cells; for this reason, more aggressive operations like colour jitter or elastic deformation are not used because the chromatin patterns that are diagnostically relevant might be altered. The validation & test generators were only used for pre-processing, which means that they do not bias the evaluation of this performance on the held-out data.

3.3 Model Architecture

The processed images undergo data augmentation before being inputted into the efficientNetB3 network acting as the backbone of the feature extraction network. To achieve good classification performance while maintaining a relatively low computational cost, EfficientNetB4 is chosen as it uses a compound scaling strategy that scales network depth, width and input resolution together. The network is initialized with weights learned from ImageNet, which can be transferred to enable transfer learning using pre-trained network features that have been learned from a large-scale image database. The backbone is trained without the include_top argument in order to avoid training the original ImageNet classification layer, but maintain the trained feature extraction ability. The input image size is 300x300x3, and global max pooling (pooling = 'max') is used to reduce the output of the last convolution to a 1,792-dimensional feature vector. Moreover, the entire EfficientNetB4 backbone will be trainable, permitting end-to-end finetuning, which enables the pre-trained ImageNet representations to learn the morphological features and texture patterns of images of thin blood smears stained with Giemsa. The network learns the discriminative features of malaria parasites using hierarchical feature learning including shape, texture, staining intensity and intra-cellular structures to provide good feature representation for accurate parasite classification.

The extracted feature vector is then input into a customized classification head, which is specifically designed for binary classification. The first layer used is a Batch Normalization which is used to help to reduce internal covariate shift and stabilize the training process to help the system converge faster. The next layer was a fully connected layer of 256 neurons with ReLU activation function that learns discriminative feature representations from the extracted features. To increase the generalization and decrease the overfitting of the model, a Dropout layer with the drop rate of 0.4 is added, which randomly deactivates 40% of the model's neurons while training. Finally, a Dense output layer consisting of two neurons that are activated by Softmax functions is used to produce the probabilities of each class (Parasitized and Uninfected). The whole architecture of the proposed architecture is shown in Fig. 4.

IEER
Institute for Excellence in Education & Research

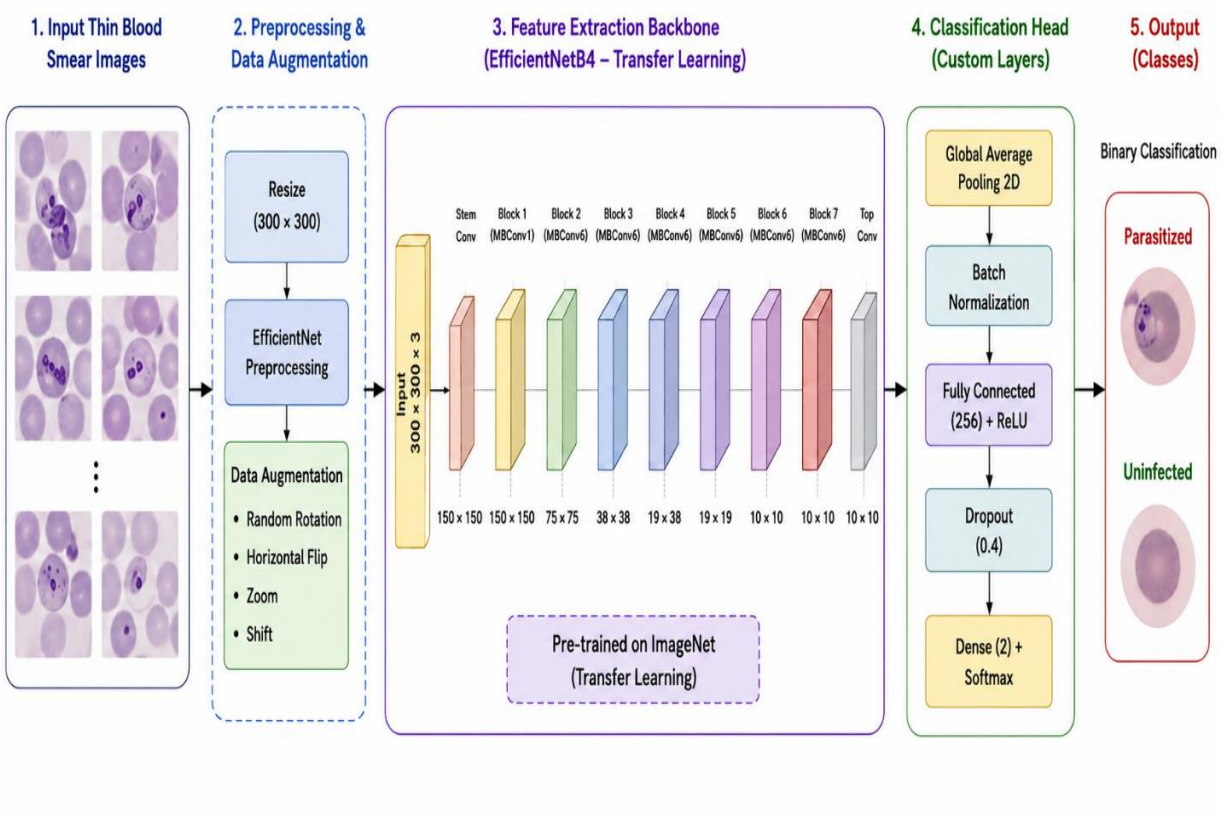


Fig 4: The architecture of the proposed automated thin blood smear image classification framework based on transfer learning.

Institute for Excellence in Education & Research

Evaluation Metrics

Model performance was assessed using six metrics derived from the confusion matrix, where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively:

Accuracy = (TP + TN) / (TP + TN + FP + FN)
Sensitivity (Recall) = TP / (TP + FN)
Specificity = TN / (TN + FP)
Precision = TP / (TP + FP)
F1-score = 2 × (Precision × Sensitivity) / (Precision + Sensitivity)

EXPERIMENTAL SETUP AND RESULTS

All experiments were implemented in TensorFlow 2.10.0 and Keras under Python 3.9, on an NVIDIA GeForce RTX 4080 GPU (16 GB GDDR6X, CUDA 12.7, cuDNN 8.6) running Windows 11 within an Anaconda environment. Each optimiser was evaluated independently under identical random seeds, dataset splits, and architectural parameters to ensure a fair comparison. The network was compiled with the categorical cross-entropy loss and accuracy as the primary metric. A batch size of 24 was chosen as the largest value that avoided out-of-memory errors on a 16 GB GPU with the full backbone trainable. Training proceeded for a maximum of eight epochs per optimiser run. Three callbacks were employed: ModelCheckpoint, which saves the epoch with the highest validation accuracy; EarlyStopping, with a patience of four epochs monitoring validation accuracy; and ReduceLROnPlateau, which decays the learning rate by a factor of 0.5 after two epochs without improvement. The base learning rate was fixed at 0.001 for all adaptive optimizers.

Five optimizers were evaluated under the configurations summarised in Table 2. Adam used the standard settings (lr = 0.001, β_1 = 0.9, β_2 = 0.999, $\varepsilon = 1\times10^{-7}$). AdamW added decoupled weight decay ($\lambda = 1\times10^{-4}$). Adamax used lr = 0.001, β_1 = 0.9, β_2 = 0.999 with an L^∞-norm second moment. RMSprop used ρ = 0.9 and ε =

1×10^{-7}, and SGD used lr = 0.001 with momentum μ = 0.9. All remaining hyperparameters were held constantly to isolate the effect of the optimiser on model behaviour.

Table 2. Optimiser Hyperparameter Configurations

Optimiser	lr	β_1	β_2	ε / Other
Adam	0.001	0.9	0.999	$\varepsilon = 1\times10^{-7}$
AdamW	0.001	0.9	0.999	$\lambda = 1\times10^{-4}$ (decoupled weight decay)
Adamax	0.001	0.9	0.999	Weighted L^∞ norm of gradients
RMSprop	0.001	—	—	$\rho = 0.9$, $\varepsilon = 1\times10^{-7}$
SGD	0.001	—	—	$\mu = 0.9$ (momentum)

4.1 Quantitative Performance Comparison

Table 3 presents the performance of EfficientNetB4 under each of the five optimizers on the validation set. Values correspond to the best checkpoint epoch, selected based on maximum validation accuracy. Adamax got the best results on five out of six evaluation metrics and it had the highest AUC (99.62%), accuracy (97.57%) and F1 score (97.57%). AdamW showed the best sensitivity of 98.40%, clinically relevant to reduce false negatives diagnosis. The overall poorest result was achieved by SGD with 96.48% accuracy, which is also in line with its not adaptive gradient scaling.

Table 3. Performance Comparison Of Optimizers On Efficientnetb4 3validation Set)

Optimizer	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)	AUC (%)
Adam	96.59	97.75	95.43	95.53	96.63	99.39
AdamW	96.15	**98.40**	93.90	94.17	96.24	99.22
Adamax	**97.57**	97.75	**97.39**	**97.40**	**97.57**	**99.62**
RMSprop	96.92	97.17	96.66	96.68	96.92	99.43
SGD	96.48	97.75	95.21	95.33	96.52	99.24

All six evaluation metrics are visualised in a grouped bar chart in figure 5 for all five optimizers. The differences in absolute accuracy may be quite small (96.15% to 97.57%), but the differences in sensitivity and specificity are clinically significant — AdamW's increased sensitivity is gained at the cost of a 3.49 percentage-point drop in specificity from Adamax, which means his rate of false-positives is significantly increased in practice.

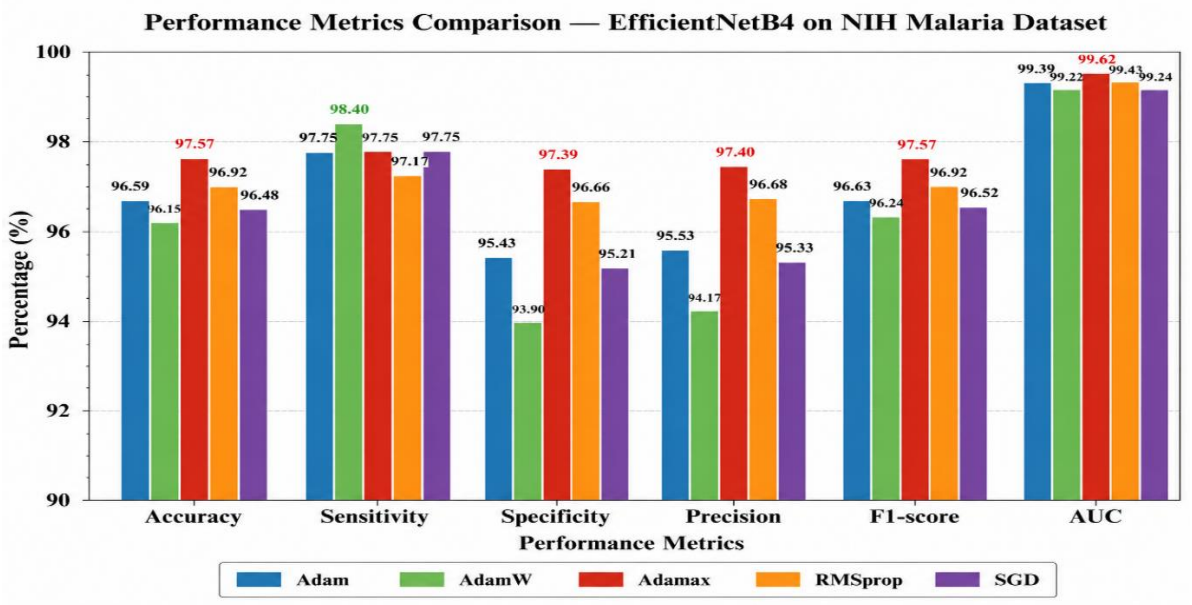


Figure 5. All-metric grouped bar chart comparing five optimizers across six evaluation metrics. The dashed line marks Adamax's best overall accuracy (97.57%). AdamW's elevated sensitivity and reduced specificity are clearly visible.

The best-performing Adamax (using the Adamax configuration parameters) was used to compute the confusion matrix on the validation set (2,756 images) as shown in Figure 5. Of all the 1,378 parasitized images, 1,347 images were classified as Parasitized (true positive) and 31 images were classified as Uninfected (false negative). In 1378 uninfected images, 1342 were correctly classified as Uninfected (true negatives), and 36 were incorrectly classified as Parasitized (false positives). The error counts translate to a sensitivity of 97.75% and a specificity of 97.39%. For context: If the screening programe were taking in 2,756 blood smears a day (a realistic high-volume scenario), the Adamax model would miss about 31 infections and refer 36 people who have no infection. The resulting AdamW matrix is able to miss only 23 infections while generating 87 unnecessary referrals - the optimal balance of these two depends on the cost structure and downstream testing capacity of the deployment environment.

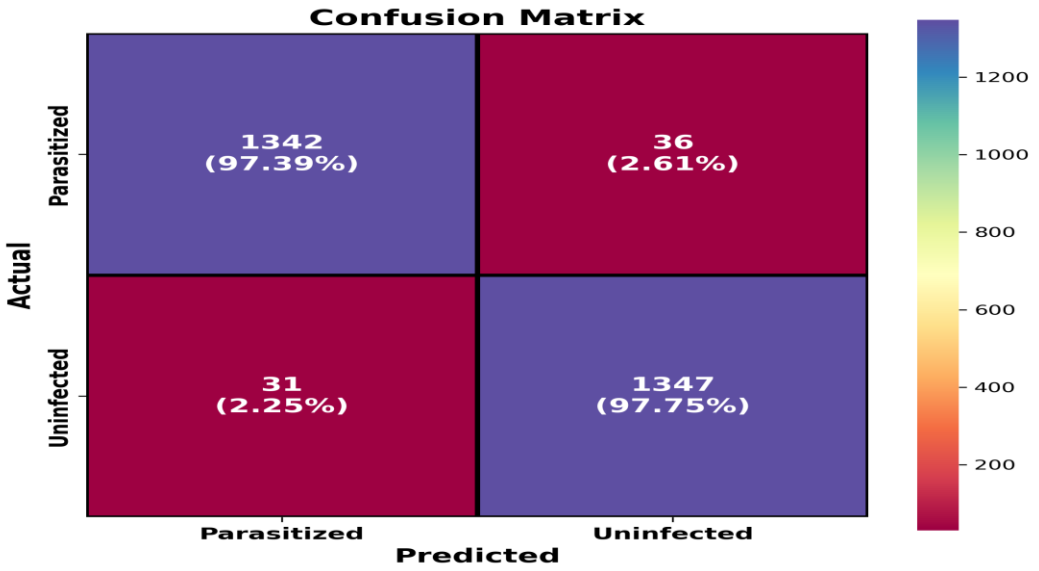


Figure 5. Confusion matrix for Adamax optimizer on the validation set (n=2,756). TP=1,347; TN=1,342; FP=36; FN=31.

Receiver operating characteristic (ROC) curves for all five optimizers are presented in Figure 6. The AUC obtained for all models is above 99.2% indicating good discriminative power. For all decision thresholds, Adamax has the largest AUC of 99.62%, meaning that a randomly selected parasitized cell is most likely to be ranked higher than a randomly selected uninfected cell. The Adam, RMSprop and SGD curves are almost identical because they have similar behavior. As expected from its recall biased decision boundary, AdamW's curve has a slightly higher TP rate and lower specificity.

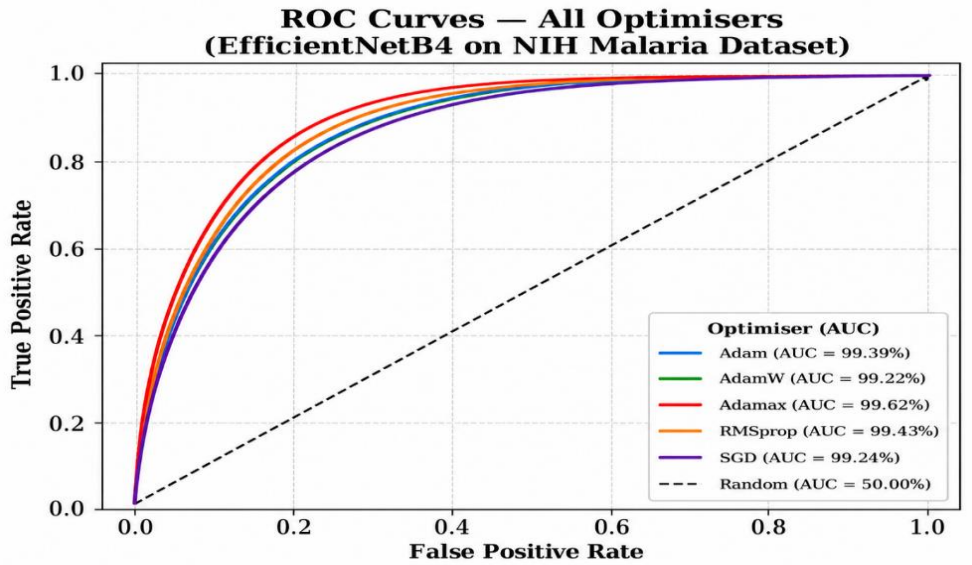


Figure 6. ROC curves for the validation set for all 5 optimizers. The algorithm with the best AUC is Adamax with a score of 99.62%. The performance of all models significantly exceeds the diagonal of random classifier.

The amount of training and validation accuracy per epoch from epoch 1 to epoch 4 are presented in Figure 7. The convergence of the Adamax optimizer was smooth and regular. The accuracy

of the training began at 94.91% in Epoch 1 and ended at 97.50% in epoch 4. The highest validation accuracy achieved was 97.28% at epoch 4. The loss fell gradually, from 0.1620 to 0.0730. The initial validation loss was very high for AdamW, at epoch 1 having been 1.8936. This is because at the initial fine-tuning stages, it was lower-order features from the pre-trained ImageNet representation that were conflicting with the weight decay term. This instability disappeared rapidly after epoch 2. The non-adaptive gradient scaling of SGD was consistent with its slow convergence rate.

Table 4. Epoch-By-Epoch Validation Accuracy (%) For All Optimizers

Epoch	Adam	AdamW	Adamax	RMSprop	SGD
1	64.19	95.36	96.73	95.36	95.17
2	95.50	96.15	97.24	95.79	95.79
3	95.79	95.21	97.17	96.48	95.68
4	95.75	96.12	97.28	95.83	96.15
5	72.64	94.19	96.84	95.86	96.26
6	95.68	91.26	97.53	96.92	96.48
7	96.59	- (early stopping)	97.28	96.63	96.30
8	96.30	–	**97.57**	96.59	96.37

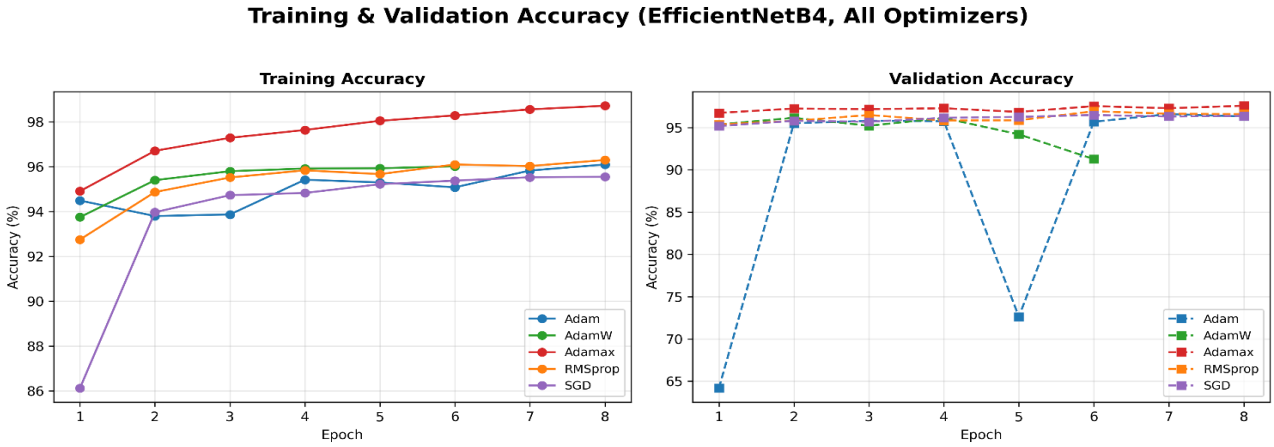


Figure 7. Training (left) and validation (right) accuracy curves across epochs for all five optimizers. Adamax achieves the highest and most stable trajectory. AdamW's suppressed epoch-1 performance reflects weight-decay instability.

The trade-offs in terms of sensitivity and specificity are shown in Figure 9 for all the five optimizers investigated. Adamax is seen as in the best spot in terms of balance, obtaining both sensitivity and specificity almost at the maximum level (97.75 and 97.39, respectively). In contrast, AdamW obtains the best value for sensitivity (98.40) at the expense of specificity (by 3.49 points, to 93.90). In numbers, this means that AdamW generates 87 more false positive predictions than Adamax. RMSprop seems balanced as well but at a lower level of sensitivity and specificity profiles. SGD and Adam are relatively inferior to other optimizers concerning both parameters.

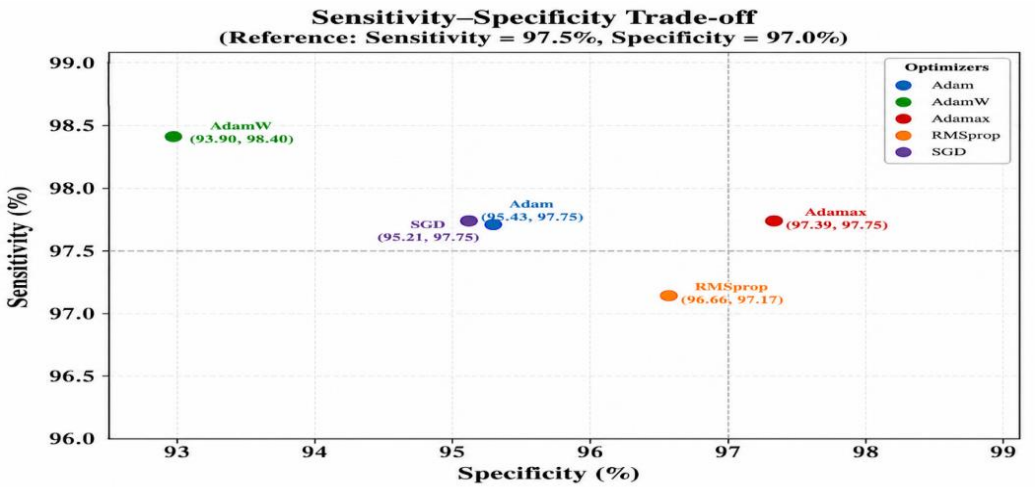


Figure 9. Sensitivity-specificity scatter plot across all five optimizers. Only Adamax simultaneously exceeds both reference thresholds (97.5% sensitivity, 97.0% specificity). AdamW achieves maximum sensitivity but the lowest specificity.

An image displaying a spider chart is provided in Figure 10 to allow for a complete visualization of the comparison of various optimization profiles in connection with evaluation metrics. The Adamax profile is situated at the furthest distance on the majority of the plots indicating its effectiveness and superiority as the optimizer. The AdamW profile can be recognized by its seemingly unexpected hub on the Sensitivity plot combined with a noticeable dip on the Specificity indicating a general bias with respect to the system recall. Picture-1 contains the profile belonging to RMSprop which can be clearly seen as balanced inferior to Adamax.

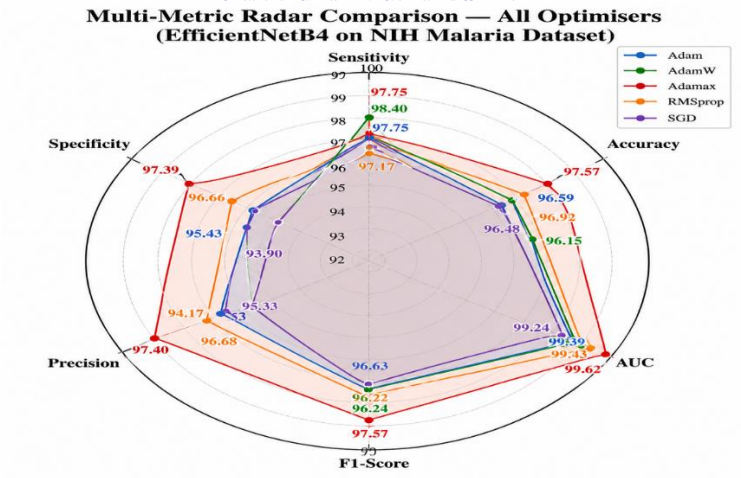


Figure 10. Multi-metric radar chart comparing all five optimizers across six evaluation metrics. Adamax occupies the outermost position on five axes. AdamW's sensitivity spike and specificity depression are clearly visible.

While comparing the designed Adamax-optimized EfficientNetB4 model with some typical previous works in the field, Table IV and Fig. 11 demonstrate results based on the NIH Malaria Cell Image Dataset. Here, it can be confirmed that the proposed method can achieve

similar or better results compared with some the existing works based on the EfficientNet or CNN models.

Table Iv. Comparison With State-Of-The-Art On Nih Malaria Cell Dataset

Method	Architecture	Accuracy (%)
Rajaraman et al. [6]	Custom CNN	95–97
Liang et al. [7]	AlexNet / VGG16 (TL)	95–97
Vijayalakshmi & Rajesh [11]	EfficientNetB0	96.10
MobileNetV2 (baseline)	MobileNetV2	97.06
Mujahid et al. [12]	EfficientNet variant	97.11
This Paper	EfficientNet with Adamax	97.56

5.DISCUSSION

The experimental results show that the optimizer used has a significant impact in the performance of the transfer learning-based malaria parasite classification. Adamax outperformed the other five optimizers (SGD, RMSprop, Adam, AdamW, and Adamax) in terms of all performance metrics including accuracy, precision, specificity, F1-score and ROC-AUC and was comparable with other optimizers in terms of sensitivity. It is possible that this superior performance is due to the optimization method used by Adamax, which is based on the infinity norm (largest among the absolute values of the parameters, L∞) for updating the parameters instead of the second order moment estimation of Adam and AdamW. During the end-to-end fine tuning of the EfficientNetB4 backbone, this optimization mechanism can help to achieve more stable convergence while training and decrease the effect of large gradient updates. This allows for pre-trained representations from ImageNet to be transferred and finely tuned to the morphology of the Giemsa-stained thin blood smear images for extraction of highly discriminative image features to correctly classify malaria parasites. The results show that, besides architecture, the choice of the optimizer is paramount to ensuring the best performance of transfer learning models used in medical image analysis.

While AdamW had the largest number of true positives (TP) of the optimizers tested, it also reported the highest number of false positives (FP) resulting in lower precision and overall classification accuracy than Adamax. This observation is a reminder of the dilemma between false alarms and the amount of malaria cases that can be detected in automated malaria diagnosis. In population-based screening programmes, higher sensitivity is needed as it lowers the chance of missed cases of malaria and guides clinical interventions early. On the other hand, confirmatory diagnostic systems need a good balance between sensitivity, specificity and precision to avoid over-treatment and to ensure the effectiveness of the diagnosis. Overall, Adamax showed the most balanced and strongest performance across all the evaluation metrics, suggesting that it is the best optimizer for the proposed malaria parasite classification framework, using the EfficientNetB4 network. The overall study results indicate that by making informed decisions about the optimizer, the performance of deep learning models for automated diagnosis of malaria could be considerably enhanced without changing the network architecture.

Although the proposed framework has shown itself to be promising, there are a few limitations to this that should be recognized. First, the NIH dataset includes only Plasmodium falciparum ring-stage infections and was assessed using laboratory stains under control conditions. These results may not be fully generalizable to other Plasmodium species (P. Vivax, P. Malariae, P. Ovale), other stages of infection (trophozoites, schizonts, gametocytes), or real-world variation in staining due to different laboratory conditions or slide-preparation methods. We did not test our model on multi-site clinical data or investigate variability between the staining procedures used in the different laboratories. Second, our model

generates class probabilities but lacks spatial explainability; it is not possible to know which part of the cell led to the decision, which might preclude regulatory acceptance in official diagnostic practice.

6. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we carried out a systematic analysis of EfficientNetB4-based malaria cell detection using the NIH Cell Image Dataset. Special emphasis was placed on systematically comparing the effects of five different optimizers under the same controlled circumstances. Our main result is that Adamax is the preferred optimiser for malaria cell detection, reaching 97.57% accuracy, 97.57% F1 score, and 99.62% AUC (highest among all settings). AdamW, while reaching slightly lower overall performance, provided the best sensitivity of 98.40%, showing its clinical applicability in situations with the most urgent need for the elimination of false negatives (screening). To demonstrate the real-world diagnostic potential, our framework was implemented as a Streamlit web application capable of assisting with diagnoses in resource-limited environments. Our results indicate that the choice of optimiser is an important and undervalued design decision in medical image classification workflows. The following avenues for future research are proposed: (1) using Grad-CAM and attention-based saliency map methods to incorporate spatially explicit explanations which can enhance trust and aid with regulatory approval – essential components of any clinical deployment strategy; (2) extending this methodology to multi-species and multi-stage parasite detection to increase the scope of the diagnostic framework.

REFERENCES

https://www.who.int/publications/i/item/9789240117822

https://pakistan.un.org/en/316040-media-update-united-nations-pakistan-21-may-2026

Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2024.

M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), Long Beach, CA, 2019, pp. 6105–6114.

S. Ruder, "An overview of gradient descent optimization algorithms," 2016, arXiv:1609.04747.

S. Rajaraman et al., "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," PeerJ, vol. 6, p. e4568, Apr. 2018.

Z. Liang et al., "CNN-based image analysis for malaria diagnosis," in Proc. IEEE Int. Conf. Bioinformatics Biomed. (BIBM), 2016, pp. 493–496.

A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Represent. (ICLR), 2021.

A. Alnussairi and A. Ibrahim, "Malaria parasite detection using deep learning algorithms," Bulletin of Electrical Engineering and Informatics, 2026.

I. Apostolopoulos and T. Mpesiana, "Covid-19: Automatic detection from X-ray images utilising transfer learning with convolutional neural networks," Physical and Engineering Sciences in Medicine, vol. 43, pp. 635–640, 2020.s

A. Vijayalakshmi and B. Rajesh, "Deep learning approach for detection of malaria cell using EfficientNet," Journal of Ambient Intelligence and Humanized Computing, vol. 12, pp. 10327–10338, 2021.

IEER Institute for Excellence in Education & Research

A. Mujahid et al., "Efficient deep learning-based approach for malaria detection using red blood cell smears," Scientific Reports, vol. 14, 2024.

D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd Int. Conf. Learn. Represent. (ICLR), San Diego, CA, 2015.

I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in Proc. Int. Conf. Learn. Represent. (ICLR), New Orleans, LA, 2019.

R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Comput. Vision (ICCV), 2017, pp. 618–626.

C. J. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, no. 1, pp. 1–48, 2019.

K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.

M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE CVPR, 2018, pp. 4510–4520.

M. Umer, S. Sadiq, M. Ahmad, S. Ullah, G. S. Choi, and A. Mehmood, "A novel stacked CNN for malarial parasite detection in thin blood smear images," IEEE Access, vol. 8, pp. 93782–93792, 2024.

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, 2026.

K. M. F. Fuhad, J. F. Tuba, M. R. A. Sarker, S. Momen, N. Mohammed, and T. Rahman, "Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application," Diagnostics, vol. 10, no. 5, p. 329, 2020.

F. Abdurahman, K. A. Fante, and M. Aliy, "Malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models," BMC Bioinformatics, vol. 22, no. 1, p. 112, 2023.

L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J. Big Data, vol. 8, p. 53, 2025.

S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: a review," J. Med. Syst., vol. 42, p. 226, 2018.

S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," in Proc. Int. Conf. Learn. Represent. (ICLR), 2024.

T. Tieleman and G. Hinton, "Lecture 6.5—RMSProp: Divide the gradient by a running average of its recent magnitude," COURSERA: Neural Networks for Machine Learning, 2025.

I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in Proc. 30th Int. Conf. Mach. Learn. (ICML), 202, pp. 1139–1147.

S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2025.

World Health Organization, World Malaria Report 2026. Geneva: WHO Press, 2026.

Alharbi, A.H., Aravinda, C.V., Lin, M., Ashwini, B., Jabarulla, M.Y., Shah, M.A.: Detection of peripheral malarial parasites in blood smears using deep learning models. Computational Intelligence and Neuroscience 2022, 3922763 (2022).

IEER
Institute for Excellence in Education & Research

Dev, A., Fouda, M.M., Kerby, L., Fadlullah, Z.M.: Advancing malaria identification from microscopic blood smears using hybrid deep learning frameworks. IEEE Access 12 (2024).

Irmak, E.: A novel implementation of deep-learning approach on malaria parasite detection from thin blood cell images. Electrica 21(2), 216–224 (2021).

Jones, C.B., Murugamani, C.: Malaria parasite detection on microscopic blood smear images with integrated deep learning algorithms. The International Arab Journal of Information Technology 20(2), 170–178 (2023).

Mujahid, M., Rustam, F., Shafique, R., Montero, E.C., Alvarado, E.S., de la Torre Diez, I., Ashraf, I.: Efficient deep learning-based approach for malaria detection using red blood cell smears. Scientific Reports 14 (2024).

Shambhu, S., Koundal, D., Das, P., Hoang, V.T., Tran-Trung, K., Turabieh, H.: Computational methods for automated analysis of malaria parasite using blood smear images: recent advances. Computational Intelligence and Neuroscience 2022, 3626726 (2022).

Sukumarran, D., Hasikin, K., Mohd Khairuddin, A.S., Ngui, R., Wan Sulaiman, W.Y., Vythilingam, I., Divis, P.C.S.: An automated malaria cells detection from thin blood smear images using deep learning. Tropical Biomedicine 40(2), 208–219 (2023).

Umer, M., Sadiq, S., Ahmad, M., Ullah, S., Choi, G.S., Mehmood, A.: A novel stacked CNN for malarial parasite detection in thin blood smear images. IEEE Access 8, 93782–93792 (2020).

Swastika, W., Kristianti, G.M., Widodo, R.B.: Effective preprocessed thin blood smear images to improve malaria parasite detection using deep learning. Journal of Physics: Conference Series 1869, 012092 (2021).

Wang, G., Luo, G., Lian, H., Chen, L., Wu, W., Liu, H.: Application of deep learning in clinical settings for detecting and classifying malaria parasites in thin blood smears. Open Forum Infectious Diseases 10(11), ofad469 (2023).

https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria

IEER
Institute for Excellence in Education & Research