JOURNAL OF MEDICO INFORMATICS

Research Paper

# Automated Malaria Diagnosis from Blood Smear Images Using Fine-Tuned VGG-19 and ResNet50 Models

Md Selim Reza[1] |

1. Department of Medicine, Henry Ford Health, Detroit, MI, United States of America

## Abstract

Accurate malaria diagnosis is essential for global health management, particularly in regions where the disease causes endemic epidemics. The conventional diagnostic approach of blood sample analysis by microscopy requires skilled assessors and is susceptible to human interpretative errors. This study employs an innovative deep learning approach that integrates the VGG-19 and ResNet50 models for the automated identification of malaria through the analysis of blood smear images. The suggested approach enhances prior research by integrating transfer learning with fine-tuning strategies, improving classification accuracy. The data collected from Kaggle is subjected to preprocessing, which includes picture scaling, noise reduction, and data augmentation to enhance generalization. Model convergence efficiency is attained using sparse categorical cross-entropy with the Adam optimizer. Early stopping prevents the model from overfitting, whereas precision, recall, F1-score, and accuracy are used for validation. VGG-19 exhibits enhanced performance after fine-tuning compared to other designs, as demonstrated by experimental data, due to its improved precision and stability. This study enhances medical AI by discovering deep learning applications for malaria detection, which reduce the need for manual diagnosis and expedite illness identification in resource-limited medical environments. Interpretability strategies enhance the clinical use of automated systems and foster confidence between healthcare practitioners and these technologies.

ISSN: 3108-2696 (Online)

## 1. Introduction

The tropical and subtropical regions experience malaria as a significant global health crisis. World Health Organization reports that yearly malaria deaths amount to hundreds of thousands, with children under five years old, along with pregnant women, showing the highest risk of the disease **(Sallam *et al.* 2025)**. Humans get this disease from *Plasmodium* parasites transmitted by female Anopheles mosquitoes during their blood-feeding activities. Personnel who treat malaria rely on swift and accurate diagnosis for implementing successful malaria interventions. The commonly utilized diagnostic methods, based on blood smear microscopic analysis and rapid diagnostic tests (RDTs), exhibit three main drawbacks: they require skilled medical staff **(Maturana *et al.* 2022)**. They are subject to subjective interpretation and human operational errors. Artificial intelligence (AI) and deep learning techniques are receiving significant attention because they enable the development of automated, efficient diagnostic tools that help address existing challenges. Modern technological advancements in computer vision, combined with deep learning approaches, enable precise medical image analysis using sophisticated models. Convolution neural networks (CNNs) represent deep learning models specifically designed for image processing tasks, demonstrating remarkable potential for medical diagnostic applications **(Chong *et al.* 2025)**. The models automatically develop the ability to differentiate between healthy and infected blood samples, requiring reduced human interaction, thereby delivering an efficient and scalable method. VGG-16 and VGG-19, as well as ResNet-34 and ResNet-50, have proven successful in medical imaging applications within the CNN design family. Transfer learning with fine-tuning enables these models to improve diagnostic precision, thereby minimizing dependence on human-operated microscopes while enhancing malaria identification in low-resource areas **(Aksoy 2024)**.

The primary objective of this study is to construct and evaluate deep learning models for malaria detection using image processing techniques. The dataset used in this study comprises high-resolution digital photographs of red blood cells sourced from Kaggle, which have been categorized as either infected or healthy **(Parveen *et al.* 2025)**. Model performance is strengthened using three preprocessing methods: scaling, contrast enhancement, and noise reduction applied to the dataset. Adopting image augmentation techniques enhances data variety and offers additional advantages for generalization **(Halloum and Ez-Zahraouy 2025)**. The study employs several CNN frameworks for training to determine which network provides the most effective method for detecting malaria cases. The evaluation of model performance includes accuracy, precision, recall, and F1-score, together with confusion matrices for a comprehensive outcomes analysis. Specific challenges must be addressed before deep learning can realize its promise as a diagnostic tool for malaria **(Ahamed *et al.* 2025)**. Labeled training data must possess both sufficient quality and quantity, as deep learning models require substantial datasets for successful learning. Modeling generalization is essential because training datasets provide models that may underperform when faced with pictures from different sources or subjected to varying staining techniques **(Karimi *et al.* 2020)**. Healthcare practitioners often exhibit skepticism towards black-box algorithms, which hinders their acceptability due to the lack of transparency in deep learning models. Medical practitioners must meticulously evaluate the possible ethical hazards, including misdiagnosis and erroneous test results. Deploying deep learning models in practical environments requires addressing technical challenges in contexts with limited computing resources. The future advancement of deep learning models for malaria detection should include ensemble methods, AI explainability methodologies, and active learning strategies to enhance their reliability and practical application **(Mennella *et al.* 2024)**. Healthcare practitioners may benefit from using deep learning technologies on smartphones to facilitate malaria diagnosis in regions with limited internet connectivity. The successful application of AI in healthcare requires policy development by medical professionals, collaboration among researchers and healthcare personnel, and the advancement of AI technology. The study develops an interpretable deep learning diagnostic tool for malaria detection, aiming to enhance early treatment opportunities and diagnostic precision for malaria patients **(Obeagu and Ogenyi 2025)**.

| | | |
|---|---|---|
| Received on | : | 2025-08-25 |
| Revised on | : | 2025-09-09 |
| Accepted on | : | 2025-09-15 |
| Published Online | : | 2026-01-28 |
| Review Model | : | Single-Blind Review |
| No. of Reviewers | : | Two |
| Edited by | : | Dr. Chandrabose Selvaraj |
| Vol and Issue | : | 02 (01) |
| Page No | : | 17-23 |
| Plagiarism Level | : | 11% and 16% (AI) |
| Correspondence | : | Md Selim Reza |
| Contact Author | : | sreza1[at]hfhs.org |

**Key Words:**
Malaria Detection
Deep Learning
Medical Image Processing
*Transfer Learning*
Healthcare
Drug design
Public Health

**DOI:** 10.64659/jomi/210735

## 2. Materials and Method

### 2.1. Dataset

This research adopted images from Kaggle, which were grouped into three sections. A model training process requires the "train" folder for images that train the model and the "val" folder for images that exist for validation checks at each epoch. The independent evaluation dataset contains unseen images in the "test" folder, which is used to measure model performance. The dataset plays a crucial role in ensuring the accuracy and reliability of the model **(Ganie *et al.* 2025)**.

### 2.2. Preprocessing of the Images

The dataset was partitioned into training, validation, and test sets with a split ratio of 70:15:15 to ensure balanced model development and evaluation. Specifically, 70 percent of the images will be utilized to train the model for the acquisition of significant feature representations, 15 percent will serve to validate the model and prevent overfitting, and the remaining 15 percent will be employed for objective testing of the model **(Sivakumar *et al.* 2024)**. The specified partitioning method is widely used in deep learning research, providing a robust framework for optimizing and assessing overall generalization. The VGG-19 and ResNet-50 models were trained on blood sample data from infected and healthy specimens. We first obtained x_train, x_test, and x_val, along with their corresponding labels, y_train, y_test, and y_val, while performing image scaling in parallel **(Sebastian and Ankayarkanni 2025)**. The defined dataset consists of x_train, which contains training image NumPy arrays, while y_train contains their corresponding labels. Additionally, x_test contains test image NumPy arrays, alongside y_test, which represents the label set. The validation images stored in NumPy arrays are displayed under the x_val label, which also stores their corresponding labels in y_val **(Mustapha *et al.* 2025)**. The dataset received its labels through our use of ImageDataGenerator. The step served to properly label data by organizing the images into categorized folders. The dataset consists of two primary categories: infected blood samples and healthy ones **(Rajpal *et al.* 2021)**.

### 2.3. Model Training

The section requires developers to construct models, compile, and train them for optimal performance. We selected VGG-19 and ResNet-50 as pre-trained models to determine whether blood samples represented infected or healthy specimens, as these models have demonstrated successful performance in image classification operations. The data lacks one-hot label encoding, so sparse categorical cross-entropy is the appropriate loss function **(Caliman Sturdza *et al.* 2025)**. This study employed sparse categorical cross-entropy as the loss function for training the model. In multi-class classification tasks, when class labels are represented as integers rather than in one-hot encoding, sparse categorical cross-entropy is frequently utilized. This loss function quantifies the divergence between the model's predicted probability distribution and the actual class value, hence guiding the optimization process to minimize classification mistakes **(Debnath *et al.* 2025)**. It offers an extra advantageous effect in large-scale classification, as it is computationally efficient and eliminates the need for explicit one-hot encoding of labels. The selection of the Adam optimizer became necessary because it dynamically adjusts learning rates to achieve efficient training performance. The training phase occurred at this point as our team applied techniques to stop overfitting from developing **(Farhan *et al.* 2025)**. A model demonstrates overfitting behavior by learning dataset-specific patterns that diminish its ability to generalize. We used early stopping to control training, as it automatically terminates when the validation loss reaches a persistent peak. Training continued for 10 epochs, followed by an additional five epochs to prevent overfitting, using a patience parameter of 5 **(Almadhor *et al.* 2025)**. The training process automatically stopped once the validation loss increased for more than five epochs. Model trainability becomes possible with x_train, x_val, y_train, and y_val data obtained through the preprocessing step. Enabling the shuffle parameter prevented the model from continuously seeing the same images in successive batch presentations **(Satpathy *et al.* 2025)**.

### 2.4. Model Evaluation

The trained model was evaluated using testing on previously unseen datasets. Accuracy alone fails to determine whether a model is operational. The model performance was assessed by running a classification report and confusion matrix. These metrics enable users to evaluate model precision and recall and calculate its F1-score, thus confirming its predictive capacity **(Li *et al.* 2025; Sluijterman *et al.* 2024)**.

### 2.5. Image Processing

Techniques that operate on images aim to achieve two objectives: image enhancement and information extraction. Several different image transformations occur within this field, where resizing operations intersect with noise reduction techniques and image contrast enhancement. Digital Image Processing techniques generate enhanced images, which are crucial for successful computer vision operations. Advanced image processing methods utilize deep neural networks and machine learning models to improve image-based information extraction **(Goceri 2023)**. Machine learning methods support the evaluation and performance enhancement of image processing methods through various architectural and loss function implementations. The proposed study aims to develop a diagnostic system that detects malaria parasites present in human blood. The technique separates infected products from healthy ones through a programmed workflow that contains multiple sequential operations **(Wang *et al.* 2023)**. The system operates using Python 3.7 as its platform within JupyterLab IDE. Image preprocessing begins by adjusting the spatial resolution and then applying contrast normalization, resizing, and noise reduction steps to normalize image dimensions. Feature extraction derives pertinent characteristics, such as histograms and digital signatures, that help achieve classification **(Chiu *et al.* 2019)**. The classification part of the dataset features two categories, where parasitized conditions receive a value of zero, while healthy conditions receive a value of one. The trained model employs the gathered labels to predict new images with dependable results in detecting malaria **(Khan *et al.* 2022)**.

## 3. Results

### 3.1. Deep Learning-Enhanced Malaria Parasite Detection

Malaria transmission occurs when infected female Anopheles mosquitoes disseminate *Plasmodium* protozoa, which infiltrate red blood cells while feeding on the blood. Research estimates indicate that the *Plasmodium falciparum* parasite accounted for 99% of malaria infections in sub-Saharan Africa in 2016. Medical confirmation of malaria relies on identifying the specific parasite responsible for the sickness. The evaluation of *P. falciparum* life cycle patterns in relation to illness progression severity requires meticulous monitoring for medical purposes. The assessment of parasite quantities enables physicians to diagnose infections, determine their severity, monitor patient recovery, evaluate therapeutic effectiveness, and identify the emergence of drug resistance. A deep learning methodology is integrated into the existing system; however, the performance of image analysis results in detecting inaccuracies due to inadequate efficiency levels. Image-based malaria detection has not attained complete success.

The overall procedure adopted in the current study is shown in **Figure 1**. The suggested approach encounters significant challenges, as creating a single way to address dataset heterogeneity is more complex than formulating algorithms tailored to specific picture types. An essential strategy must be formulated to address data heterogeneity or to preprocess the information into a consistent structure for efficient execution. The primary dataset for this study comprises high-resolution digital photographs sourced from a Kaggle database. Studying red blood cells using thin blood smears is the foundation for detecting malaria parasites (MP). The integrated system employs processing technologies and machine learning techniques to enhance detection outcomes. The image processing techniques prepare the pictures before analysis by machine learning algorithms to ascertain the presence of malaria parasites in cells. Model performance enhancement necessitated the implementation of Convolutional Neural Network (CNN) models via transfer learning, supplemented with additional dataset entries for training. The fine-tuning technique initially involved freezing model layers and then refining them. The documented accuracy data yielded findings before the fine-tuning of characteristics, enabling a comparison analysis. The models underwent evaluation via the development of confusion matrices for performance assessment.

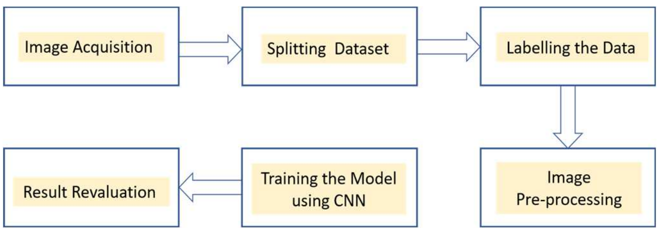


**Figure 1.** The overall methodology adopted in the current study is depicted.

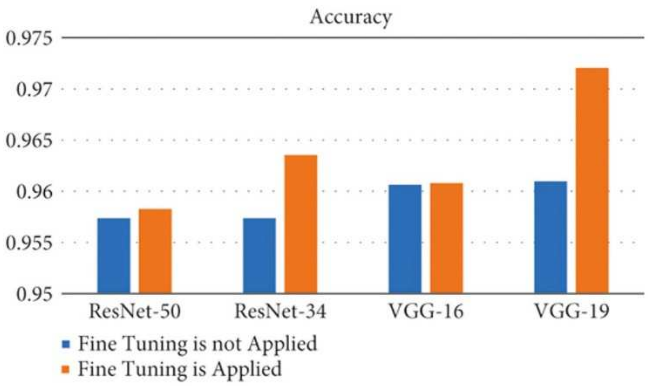


**Figure 2.** The bar chart compares the accuracy of four CNN models—ResNet-50, ResNet-34, VGG-16, and VGG-19—before (blue) and after (orange) fine-tuning.

## 3.2. Fine-tuning model performance

The **figure 2** illustrates the accuracy evaluation of the ResNet-50 and ResNet-34 models alongside the VGG-16 and VGG-19 models, both with and without fine-tuning. The picture displays models on the x-axis and accuracy scores ranging from 0.95 to 0.975 on the y-axis. The models have two bars, with blue indicating accuracy before fine-tuning and orange denoting accuracy after fine-tuning. The data shows that fine-tuning increases accuracy across all evaluated models; however, the extent of improvement varies, as seen by the data points. The accuracy of ResNet-50 is nearly identical across fine-tuned and non-fine-tuned versions, indicating that the model performs well based on its fundamental architecture. The accuracy of ResNet-34 improves significantly after fine-tuning, suggesting that further training is necessary to achieve its optimal performance. ResNet-34 relies on fine-tuning to achieve optimal performance, as it excels with supplementary training modifications.

The VGG-16 model exhibits somewhat enhanced performance after fine-tuning, adapting well without necessitating further parameter alterations. After its fine-tuning deployment, the VGG-19 model yields superior test outcomes. VGG-19's proficiency in learning via fine-tuning arises from its substantial number of parameters and superior feature extraction skills.

According to the evaluated accuracy criteria, VGG-19 demonstrates superior capabilities for malaria detection, justifying its selection as the major model for this research project. Fine-tuning enhances the accuracy of ResNet models; nevertheless, the degree of improvement between ResNet-34 and ResNet-50 is not uniform. The VGG models exhibit consistent performance, with VGG-19 achieving peak accuracy after fine-tuning, which validates its efficacy. The study emphasizes that model fine-tuning is essential for enhancing performance since deeper networks like VGG-19 provide optimal accuracy in malaria diagnosis.

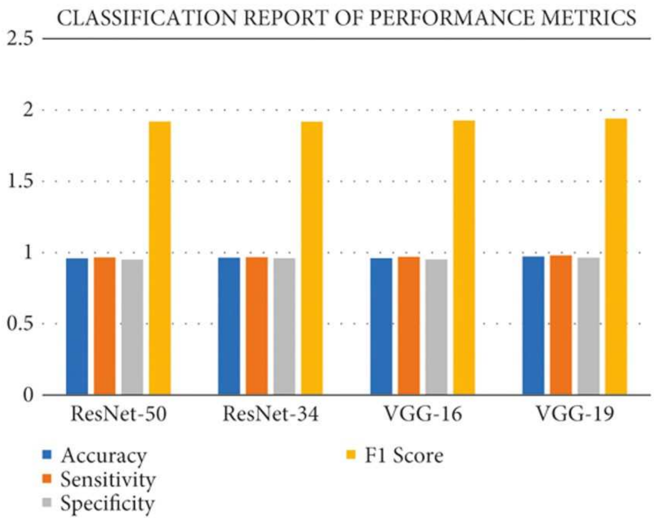


**Figure 3.** Classification report comparing the performance metrics of ResNet-50, ResNet-34, VGG-16, and VGG-19 based on accuracy, sensitivity, specificity, and F1 score. VGG-19 demonstrates the highest overall performance, particularly in terms of F1 score, indicating its effectiveness for malaria detection.

## 3.3. Classification report

The **figure 3** presents a classification report detailing the performance characteristics of the CNN models ResNet-50, ResNet-34, VGG-16, and VGG-19. The image illustrates the accuracy, sensitivity, specificity, and F1 score metrics of various models. The study delineates four distinct colour segments for each model: blue segments represent the accuracy, orange segments denote sensitivity, grey segments illustrate specificity, and yellow segments signify the F1 score. The metrics range from 0 to 2.5 on the vertical axis, while the horizontal axis displays several model names. All four designated models exhibit comparable accuracy levels, with analogous sensitivity and specificity metrics, as seen in the graph. The categorization metrics of these models exhibit alignment in the **figure 3**, as the bar groupings remain closely positioned for each model assessment. VGG-19 has relatively higher accuracy and sensitivity rates than the other evaluated models, suggesting its enhanced capability for identifying malaria-infected samples. The performance of ResNet-50 and ResNet-34 is comparable to that of the VGG models, although with a slightly reduced sensitivity rate. The models get satisfactory overall classification outcomes; nonetheless, they sporadically misidentify some positive instances. The image illustrates the exceptional F1 score, shown by yellow bars, which exceeds the measurements of other metrics across all four analyzed models. The F1 score is a crucial performance indicator, as it integrates accuracy in classification with retrieval efficacy for medical diagnoses, where incorrect identifications may result in severe consequences. All models yield significant F1 scores, indicating high accuracy and memory capabilities, which facilitate reliable malaria case identification. The same tendencies shown by VGG-16 indicate that VGG structures with increased depth have superior potential for malaria classification applications. Despite their competent performance, the F1 score of ResNet models does not exhibit the same degree of growth as that of VGG models. The integration of elevated operational velocity and exceptional precision distinguishes VGG-19 as the preeminent model among ResNet and VGG architectures.

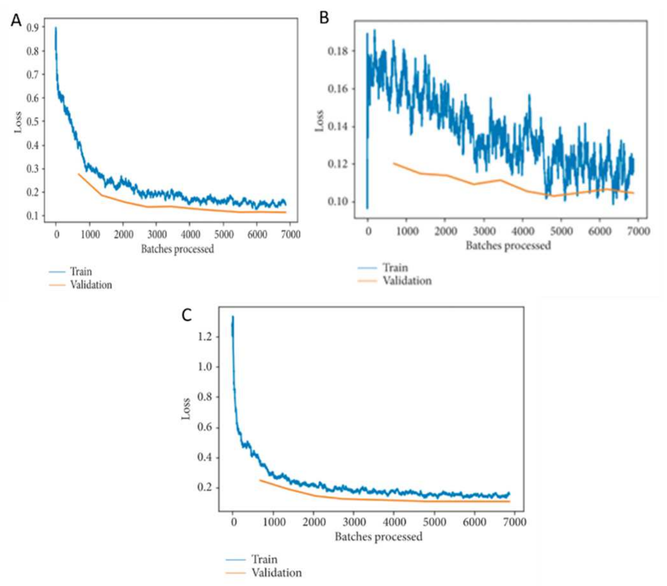


**Figure 4.** Tuned training and validation loss curves for ResNet-50 (A), ResNet-34 (B), and VGG-16 (C) models. The graphs illustrate the decrease in loss over training batches, with ResNet-34 showing fluctuations in validation loss, while VGG-16 exhibits the most stable performance.

### 3.4. Model Validation

The graph in **Figure 4** shows the progress of training and validation loss curves over time for three deep learning models: ResNet-50 (A), ResNet-34 (B), and VGG-16 (C). Each graph uses batch processing numbers as values along the x-axis and shows loss value on the y-axis. Deep learning models heavily depend on loss functions to determine prediction accuracy against actual label assessments. The performance level of a model increases as loss values decrease. Two curves are present in each graph, representing training loss (blue) and validation loss (orange). The loss metrics in subplot (A) representing ResNet-50 show continuous minimization throughout the training of multiple batches. The model progressively learns new concepts, as evidenced by a consistently downward trend in the training loss. The validation loss shows a parallel decrease but reaches a stable point, indicating proper model generalization that avoids major overfitting problems. ResNet-50 successfully extracts pertinent data features, enabling it to function with generalized performance.

The second subplot (B) shows the loss evolution of the ResNet-34 architecture. The training loss declines steadily, similar to ResNet-50, although the validation loss exhibits discontinuous variations. The model demonstrates limited stability in validation processes because it reacts sharply to changes in dataset contents. The validation loss from ResNet-34 demonstrates overall reductions; however, its fluctuating pattern indicates that the model requires further fine-tuning alongside regularization techniques to achieve stability. The training and validation loss evolution of the VGG-16 model appears in subplot (C). The training loss exhibits an initial steep descent pattern, indicating that the learning process is effective. The VGG-16 model demonstrates stable performance in its validation loss data, which decreases uniformly. VGG-16 delivers the most reliable validation loss performance compared to the other models, thus demonstrating its potential as the most dependable model for training and validation datasets. The presented **figure 4** provides valuable insights into the training processes for each model. The training and validation loss records exhibit a decreasing pattern across all models, with VGG-16 demonstrating the best stability in validation results compared to ResNet-34 and the others. ResNet-50 strikes a balance between practical learning and generalization. Malaria detection methods using deep learning become more effective when researchers select models and perform tuning, due to the architectural differences that affect efficiency and stability.

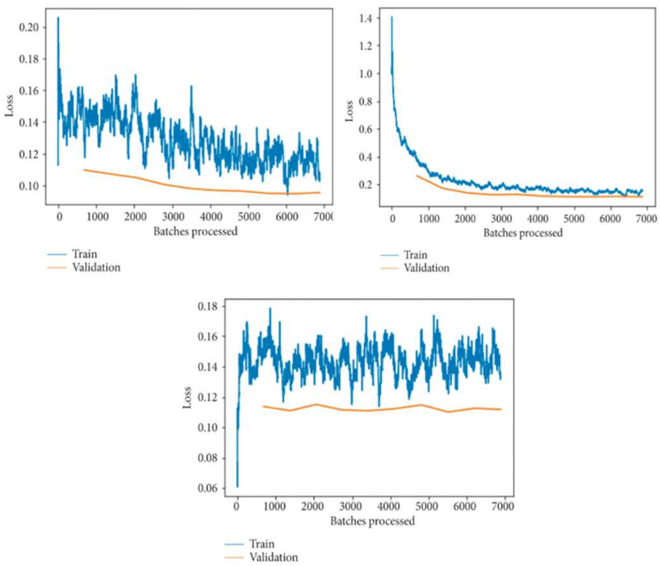


**Figure 5.** VGG-16 Model Loss Function: (A) Training and validation loss over batches processed, (B) Exponential decrease in training loss, and (C) Comparison of training and validation loss trends.

### 3.5. Validation Loss Function of VGG-16 Model

The **figure 5** illustrates the loss function behavior of the VGG-16 model, showing loss variation during training and validation across various batch sizes. The training and validation loss curves in subfigure (A) show the model's learning progress. The training and validation loss exhibit distinct patterns, with the training loss displaying fluctuations while the validation loss remains stable, indicating little overfitting. Deep learning models often show a slight disparity between training and validation loss measures. Subfigure (B) illustrates that the model exhibits rapid weight modification, leading to a swift reduction in error. The quantitative loss metric decreases exponentially throughout the training process until it attains its smallest stable value, indicating convergence. The consistent decline in validation loss suggests the model's proficiency in generalizing to new, previously unseen data points. These training patterns exemplify conventional deep learning techniques, whereby early epochs provide significant model development, followed by performance optimization with minor adjustments to model weights. The curve patterns in Subfigure (C) elucidate other aspects of training and validation loss behaviors. The model exhibits variable performance during training, as indicated by the fluctuating training loss scores, whereas the validation loss remains consistent. The disparity between the training and validation curves suggests a modest overfitting scenario, where the model exhibits exceptional performance on the training data but shows only marginal improvements on the validation data. The model maintains satisfactory generalization, as indicated by its validation loss, which exhibits negligible increases. Enhanced stability in the model's performance can be achieved by employing training loss stabilization approaches, such as regularization in conjunction with dropout or fine-tuning procedures. The loss curves shown in the picture demonstrate that the VGG-16 model effectively assimilates data while minimizing mistakes during training. The model exhibits strong generalization capabilities, as shown by consistent validation loss results; however, further modifications may further enhance its performance. Analyzing the progression of the loss function during training is crucial for effective model learning, as it helps prevent overfitting.

## 4. Discussion

The study findings indicate that deep learning methodologies, namely VGG-16 and VGG-19, exhibit superior efficacy in detecting malaria using visual data categorization approaches. The dataset acquired from Kaggle underwent preprocessing, which enhanced picture quality and facilitated fast training processes **(Jameela *et al.* 2022)**. The categorization of blood samples into healthy and diseased categories is achieved using pre-trained VGG-19 in conjunction with ResNet50, with sparse categorical cross-entropy as the loss function and the Adam optimizer to enhance training efficiency. The early halting strategy prevented the models from overfitting, preserving their powerful generalization capability. The analysis shown in **Figure 1** demonstrated that fine-tuning improved the test accuracy of each CNN model examined in the research. The optimized VGG-19 exhibited peak accuracy, demonstrating its robust ability to extract intricate characteristics from the given dataset **(Ur Rehman *et al.* 2025)**. Fine-tuning resulted in superior performance for ResNet-34 compared to its counterpart, ResNet-50. VGG-19 exhibited the most significant improvement from parameter changes, as deeper network architectures tend to gain more advantages from such tweaks. **Figure 2** presents the categorization report that validates the performance of several models. The classification parameters, including accuracy and specificity, were consistent across all models; nevertheless, VGG-19 attained the highest F1 score. The exceptional performance of VGG-19 stems from its ideal precision-recall equilibrium, making it an optimum instrument for malaria diagnosis. The detection sensitivity of ResNet models was marginally lower than that of VGG-based models, which may compromise their ability to accurately identify infected individuals **(Kandhro *et al.* 2024)**. **Figure 3** illustrates the validation loss curves, which facilitated the assessment of training stability levels. The validation loss of VGG-16 remained stable throughout training, indicating robust generalization stability. ResNet-50 exhibited a stable learning pattern, but ResNet-34 required more parameter modifications to attain system stability. The examination of the VGG-16 loss function revealed steady validation loss features despite modest oscillations in training loss, as it effectively mitigated the danger of overfitting. In our experiment, VGG-19 consistently outperformed ResNet50 and the other models tested. A variety of variables can be ascribed to this performance disparity. The data set utilized in this study may exhibit bias towards models featuring deeper convolutional layers and larger receptive fields, which allow VGG-19 to discern finer-grained spatial information. Secondly, the fine-tuning method employed in our work likely enabled VGG-19 to utilize the augmented number of parameters without significant overfitting, hence improving its feature discrimination. Conversely, although the residual learning architecture of ResNet50 is highly efficient and mitigates vanishing gradients, it may also lead to representational smoothing, which is less suitable for the intricate details in our dataset. Ultimately, architectural inductive biases, such as the uniform layering of VGG-19 and its basic design, may align more effectively with the underlying data structure than the intricate residual blocks of ResNet50 **(Afifi *et al.* 2025)**. All of these factors may contribute to the efficacy of VGG-19 in this context, despite it being an older and more parameter-intensive model.

### 4.1. Challenges

Acquiring high-quality, labeled training data for deep learning models is challenging due to the significant barrier posed by the expert annotation of infected and uninfected cells. Creating a substantial and diversified database is a prerequisite for constructing successful models. The generalization capacity of trained models is limited due to their restricted adaptability to diverse populations and variations in staining techniques. The ability to comprehend model decision-making is essential, as deep learning systems often operate as black boxes, complicating clinicians' understanding of their processes **(Dong *et al.* 2025)**. Medical treatment may have significant repercussions owing to ethical hazards arising from possible misdiagnoses and erroneous test findings. Implementing these models in resource-constrained regions faces two primary operational obstacles: limitations on computer power and a consistent electricity supply **(Goktas and Grzybowski 2025)**. Ongoing education is crucial, as the malaria parasite may mutate and new strains and imaging techniques may emerge; therefore, the model requires regular modifications to sustain accuracy and efficacy over time.

### 4.2. Future Directions

Combining multiple deep learning models or deep learning models integrated with traditional image processing techniques through ensemble methods enhances detection accuracy and robustness in the malaria diagnostic process. Developing explainable AI methods represents a critical need to interpret model decisions to gain the trust of medical experts and regulatory bodies. Active learning systems that empower human oversight allow models to study uncertain cases without requiring large amounts of expert tagging **(Awe *et al.* 2025)**. The availability of edge computing with analytical models installed on mobile smartphones enables medical staff to diagnose malaria in remote areas lacking internet connectivity. The successful application of computer models can only be achieved through ongoing partnership work between scientists from medical fields and computer science researchers who will tackle complicated problems and enhance the models. Through image processing technologies augmented with deep learning methods, healthcare providers can refine diagnosis processes and achieve better medical outcomes, particularly in endemic areas **(Alrashdi 2024)**. Beyond its usability, researchers must carefully examine the system's restrictions, both ethical aspects and steady evaluation procedures, to maintain its dependability and operational capability. In the future, various solutions may be explored to address the ethical issues and deployment limitations of deep learning models in practice. Pruning and quantization are model compression techniques that significantly reduce computational requirements and memory use, facilitating deployment on resource-constrained devices with minimal performance degradation. Federated learning offers a significant possibility to enhance data privacy and security through the decentralized training of models across several institutions without the need to share actual data **(Mennella *et al.* 2024)**. These strategies, combined with the development of fairness-oriented algorithms and explainable AI, offer concrete pathways to overcome current limitations and enhance the responsibility, efficiency, and equity of AI models in practice.

## 5. Conclusion

This research concludes that medical image analysis requires selecting appropriate models and precisely tuning them to achieve optimal results. The VGG-19 model became the best choice because it demonstrated the highest levels of accuracy alongside a superior F1 score and generalization competence. Deep learning systems integrated with image processing methods can enhance the automation of malaria detection, thereby building more reliable and efficient diagnosis strategies. The research should focus on advanced model optimization using expanded data collections and optimized techniques to enhance classification precision.

## 6. Disclosure Statements

### 6.1. Author Contribution

**MSR:** Conceptualization, Methodology, Model Development, Data Analysis, Writing—Original Draft Preparation, and Writing—Review & Editing. **MSR** have read and approved the final manuscript.

## 6.2. Declaration of Generative AI

The authors declare that no generative AI tools were used in the drafting, writing, or editing of the manuscript. All scientific interpretations and conclusions are the author's own.

## 6.3. Ethics approval (for clinical/animal studies)

Ethical review and approval were waived for this study as it did not involve human participants or animals. The data used was obtained from an open-access dataset (Kaggle) that is publicly available.

## 6.4. Informed Consent Statement

Not applicable.

## 6.5. Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## 6.6. Acknowledgment

The author acknowledges Henry Ford Health, Department of Medicine, Detroit, MI, USA, for institutional support.

## 6.7. Funding Statement

No funding was reported for this research work

## 6.8. Conflicts of Interest

The authors declare that they have no known financial, personal, academic, or other relationships that could inappropriately influence, or be perceived to influence, the work reported in this manuscript. The author confirms that there are no competing interests to declare.

## 6.9. Corresponding Author Contact Information

The corresponding author **Dr. Md Selim Reza** can be contacted via email sreza1[at]hfhs.org.

## 7. Reference


Afifi I, Elgendy M, Abdelfatah M, El-Sappagh S. (2025), Vision and convolutional transformers for Alzheimer's disease diagnosis: a systematic review of architectures, multimodal fusion and critical gaps, Brain Inform, 13(1):1. **doi:**10.1186/s40708-025-00286-7. **PMID:** 41405782.

Ahamed MF, Nahiduzzaman M, Mahmud G, Shafi FB, Ayari MA, Khandakar A, Abdullah-Al-Wadud M, Islam SMR. (2025), Improving Malaria diagnosis through interpretable customized CNNs architectures, Sci Rep, 15(1):6484. **doi:**10.1038/s41598-025-90851-1. **PMID:** 39987229.

Aksoy A. (2024), An Innovative Hybrid Model for Automatic Detection of White Blood Cells in Clinical Laboratories, Diagnostics (Basel), 14(18) **doi:**10.3390/diagnostics14182093. **PMID:** 39335772.

Almadhor A, Alsubai S, Kryvinska N, Hejaili AA, Ayari M, Bouallegue B, Abbas S. (2025), Evaluating large transformer models for anomaly detection of resource-constrained IoT devices for intrusion detection system, Sci Rep, 15(1):37972. **doi:**10.1038/s41598-025-21826-5. **PMID:** 41168224.

Alrashdi I. (2024), Fog-based deep learning framework for real-time pandemic screening in smart cities from multi-site tomographies, BMC Med Imaging, 24(1):123. **doi:**10.1186/s12880-024-01302-8. **PMID:** 38797827.

Awe OO, Mwangi PN, Goudoungou SK, Esho RV, Oyejide OS. (2025), Explainable AI for enhanced accuracy in malaria diagnosis using ensemble machine learning models, BMC Med Inform Decis Mak, 25(1):162. **doi:**10.1186/s12911-025-02874-3. **PMID:** 40217281.

Caliman Sturdza OA, Filip F, Terteliu Baitan M, Dimian M. (2025), Deep Learning Network Selection and Optimized Information Fusion for Enhanced COVID-19 Detection: A Literature Review, Diagnostics (Basel), 15(14) **doi:**10.3390/diagnostics15141830. **PMID:** 40722579.

Chiu YC, Chen HH, Zhang T, Zhang S, Gorthi A, Wang LJ, Huang Y, Chen Y. (2019), Predicting drug response of tumors from integrated genomic profiles by deep neural networks, BMC Med Genomics, 12(Suppl 1):18. **doi:**10.1186/s12920-018-0460-9. **PMID:** 30704458.

Chong PL, Vaigeshwari V, Mohammed Reyasudin BK, Noor Hidayah BRA, Tatchanaamoorti P, Yeow JA, Kong FY. (2025), Integrating artificial intelligence in healthcare: applications, challenges, and future directions, Future Sci OA, 11(1):2527505. **doi:**10.1080/20565623.2025.2527505. **PMID:** 40616302.

Debnath K, Rana P, Ghosh P. (2025), A survey on deep learning for drug-target binding prediction: models, benchmarks, evaluation, and case studies, Brief Bioinform, 26(5) **doi:**10.1093/bib/bbaf491. **PMID:** 40977267.

Dong C, Liu Y, Nie J, Zhang X, Yu F, Zhou Y. (2025), Artificial Intelligence in Infectious Disease Diagnostic Technologies, Diagnostics (Basel), 15(20) **doi:**10.3390/diagnostics15202602. **PMID:** 41153274.

Farhan M, Waheed Ud Din H, Ullah S, Hussain MS, Khan MA, Mazhar T, Khattak UF, Jaghdam IH. (2025), Network-based intrusion detection using deep learning technique, Sci Rep, 15(1):25550. **doi:**10.1038/s41598-025-08770-0. **PMID:** 40664956.

Ganie SM, Malik MB, Aadil M, Muteeb G, Farhan M, Aatif M. (2025), Explainable AI based hybrid DRM-Net transfer learning model for breast cancer detection and classification using ultrasound images, Sci Rep, 15(1):44170. **doi:**10.1038/s41598-025-27934-6. **PMID:** 41419544.

Goceri E. (2023), Medical image data augmentation: techniques, comparisons and interpretations, Artif Intell Rev, 10.1007/s10462-023-10453-z:1-45. **doi:**10.1007/s10462-023-10453-z. **PMID:** 37362888.

Goktas P, Grzybowski A. (2025), Shaping the Future of Healthcare: Ethical Clinical Challenges and Pathways to Trustworthy AI, J Clin Med, 14(5) **doi:**10.3390/jcm14051605. **PMID:** 40095575.

Halloum K, Ez-Zahraouy H. (2025), Enhancing Medical Image Classification through Transfer Learning and CLAHE Optimization, Curr Med Imaging, 21:e15734056342623. **doi:**10.2174/0115734056342623241119061744. **PMID:** 40259867.

Jameela T, Athotha K, Singh N, Gunjan VK, Kahali S. (2022), Deep Learning and Transfer Learning for Malaria Detection, Comput Intell Neurosci, 2022:2221728. **doi:**10.1155/2022/2221728. **PMID:** 35814548.

Kandhro IA, Manickam S, Fatima K, Uddin M, Malik U, Naz A, Dandoush A. (2024), Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification, Heliyon, 10(10):e31488. **doi:**10.1016/j.heliyon.2024.e31488. **PMID:** 38826726.

Karimi D, Dou H, Warfield SK, Gholipour A. (2020), Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis, Med Image Anal, 65:101759. **doi:**10.1016/j.media.2020.101759. **PMID:** 32623277.

Khan SH, Shah NS, Nuzhat R, Majid A, Alquhayz H, Khan A. (2022), Malaria parasite classification framework using a novel channel squeezed and boosted CNN, Microscopy (Oxf), 71(5):271-282. **doi:**10.1093/jmicro/dfac027. **PMID:** 35640304.

Li C, Dai J, Tian S, Jiang Y, Hu Q, Chen J, Peng S, Wen Y. (2025), Ranking-oriented machine learning framework for probabilistic wind power forecasting with temporal reliability constraints, Sci Rep, 15(1):42125. **doi:**10.1038/s41598-025-26241-4. **PMID:** 41298734.

Maturana CR, de Oliveira AD, Nadal S, Bilalli B, Serrat FZ, Soley ME, Igual ES, Bosch M, Lluch AV, Abello A, Lopez-Codina D, Sune TP, Clols ES, Joseph-Munne J. (2022), Advances and challenges in automated malaria diagnosis using digital microscopy imaging with artificial intelligence tools: A review, Front Microbiol, 13:1006659. **doi:**10.3389/fmicb.2022.1006659. **PMID:** 36458185.

Mennella C, Maniscalco U, De Pietro G, Esposito M. (2024), Ethical and regulatory challenges of AI technologies in healthcare: A narrative review, Heliyon, 10(4):e26297. **doi:**10.1016/j.heliyon.2024.e26297. **PMID:** 38384518.

Mustapha B, Zhou Y, Shan C, Xiao Z. (2025), Enhanced Pneumonia Detection in Chest X-Rays Using Hybrid Convolutional and Vision Transformer Networks, Curr Med Imaging, 21:e15734056326685. **doi:**10.2174/0115734056326685250101113959. **PMID:** 39806960.

Obeagu EI, Ogenyi FC. (2025), Artificial intelligence in African malaria control programs: opportunities and risks, Ann Med Surg (Lond), 87(12):8664-8670. **doi:**10.1097/MS9.0000000000004259. **PMID:** 41377369.

Parveen R, Qui B, Song W, Al-Kahtani N, Jamjoom MM, Mostafa SM, Sultan N, Fatima J. (2025), Trustworthy deep learning for malaria diagnosis using explainable artificial intelligence, Sci Rep, 15(1):45037. **doi:**10.1038/s41598-025-28387-7. **PMID:** 41419508.

Rajpal S, Lakhyani N, Singh AK, Kohli R, Kumar N. (2021), Using handpicked features in conjunction with ResNet-50 for improved detection of COVID-19 from chest X-ray images, Chaos Solitons Fractals, 145:110749. **doi:**10.1016/j.chaos.2021.110749. **PMID:** 33589854.

Sallam M, Al-Khatib AO, Al-Mahzoum KS, Abdelaziz DH, Sallam M. (2025), Current Developments in Malaria Vaccination: A Concise Review on Implementation, Challenges, and Future Directions, Clin Pharmacol, 17:29-47. **doi:**10.2147/CPAA.S513282. **PMID:** 40191019.

Satpathy S, Tripathy U, Swain PK. (2025), Cloud-based DDoS detection using hybrid feature selection with deep reinforcement learning (DRL), Sci Rep, 15(1):36546. **doi:**10.1038/s41598-025-18857-3. **PMID:** 41120397.

Sebastian N, Ankayarkanni B. (2025), Enhanced ResNet-50 with Multi-Feature Fusion for Robust Detection of Pneumonia in Chest X-Ray Images, Diagnostics (Basel), 15(16) **doi:**10.3390/diagnostics15162041. **PMID:** 40870892.

Sivakumar M, Parthasarathy S, Padmapriya T. (2024), Trade-off between training and testing ratio in machine learning for medical image processing, PeerJ Comput Sci, 10:e2245. **doi:**10.7717/peerj-cs.2245. **PMID:** 39314694.

Sluijterman L, Cator E, Heskes T. (2024), How to evaluate uncertainty estimates in machine learning for regression?, Neural Netw, 173:106203. **doi:**10.1016/j.neunet.2024.106203. **PMID:** 38442649.

Ur Rehman S, Damasevicius R, Al Sukhni H, Aljohani A, Hamza A, Mohammed Alsekait D, AbdElminaam DS. (2025), A novel deep learning based approach with hyperparameter selection using grey wolf optimization for leukemia classification and hematologic malignancy detection, PeerJ Comput Sci, 11:e3160. **doi:**10.7717/peerj-cs.3160. **PMID:** 40989368.

Wang G, Luo G, Lian H, Chen L, Wu W, Liu H. (2023), Application of Deep Learning in Clinical Settings for Detecting and Classifying Malaria Parasites in Thin Blood Smears, Open Forum Infect Dis, 10(11):ofad469. **doi:**10.1093/ofid/ofad469. **PMID:** 37937045.



**How to Cite:** Reza, M.S., (2026). Automated Malaria Diagnosis from Blood Smear Images Using Fine-Tuned VGG-19 and ResNet50 Models. *Journal of Medico Informatics*, 2(1), 17-23. **doi:** http://doi.org/10.64659/jomi/210735