Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 6, December 2025, pp. 4590~4600
ISSN: 2302-9285, DOI: 10.11591/eei.v14i6.10158

Meta-learning for malaria diagnosis: evaluating stacking models for enhanced classification performance

Komal Kumar Napa[1], Sangeetha Murugan[2], Sathya Subramanian[3], Durga Devi Saravanan[4], Devana Nageswari[5], Battula Krishna Prasad[6]
[1]Department of Artificial Intelligence and Data Science, Saveetha Engineering College, Chennai, India
[2]Department of Computer Science & Engineering, Madanapalle Institute of Technology & Science (MITS), Deemed to be University, Madanapalle, India
[3]Department of Electronics and Communication Engineering, S. A. Engineering College, Chennai, India
[4]Department of Computer Science and Engineering, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India
[5]Department of Science and Humanities (General Engineering-EEE), R.M.K. College of Engineering and Technology, Chennai, India
[6]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

Article Info

Article history:

Received Feb 25, 2025
Revised Sep 16, 2025
Accepted Sep 27, 2025

Keywords:

Deep learning
Malaria detection
Meta-learning
Stacking ensemble
XGBoost

ABSTRACT

Accurate malaria detection is crucial for effective disease management, particularly in regions with limited medical resources. Deep learning models have shown promising results in automated diagnosis, yet real-world deployment often faces challenges such as computational cost and model interpretability. This study evaluates multiple deep learning architectures—VGG16, ResNet50, InceptionV3, MobileNetV2, and DenseNet121—on the publicly available National Institutes of Health (NIH) malaria cell image dataset (27,558 images), and enhances their performance using stacking ensemble learning with different meta-learners. Among individual models, DenseNet121 achieved the highest accuracy of 88.00%, while MobileNetV2 had the lowest at 84.80%. Implementing stacking with logistic regression as the meta-learner improved accuracy to 89.40%, while random forest increased it to 90.10%. The best performance was achieved with XGBoost as the meta-learner, attaining an accuracy of 91.20%, precision of 92.10%, recall of 90.80%, and an F1-score of 91.40%—representing a 3.2% accuracy improvement over the best individual model. The classification report further confirms superior performance in distinguishing infected and uninfected cases. These results highlight the potential of stacking with advanced meta-learners to support health workers in rapid, reliable malaria diagnosis, ultimately aiding timely treatment, and improving patient outcomes in clinical and field settings.

Corresponding Author:

Komal Kumar Napa
Department of Artificial Intelligence and Data Science, Saveetha Engineering College
Chennai, India
Email: komalkumarnapa@gmail.com

1. INTRODUCTION

Malaria remains one of the most significant global health threats, affecting millions of people annually [1]. It is primarily caused by Plasmodium parasites, which are transmitted through the bites of infected female Anopheles mosquitoes. Despite advances in medical technology, timely and accurate diagnosis remains a challenge, particularly in resource-limited regions. Traditional diagnostic techniques, such as microscopic examination of blood smears and rapid diagnostic tests (RDTs), have been widely

utilised [2]. However, these methods are often labour-intensive, subjective, and require skilled professionals for accurate interpretation. To address these limitations, automated diagnostic approaches based on deep learning have gained considerable attention in recent years [3]. The application of deep learning models to medical image analysis has demonstrated remarkable success in various domains, including the classification of pathological conditions in radiology and dermatology.

Convolutional neural networks (CNNs) have been extensively explored for image-based disease detection due to their ability to learn hierarchical features directly from raw input data. Several CNN architectures, including VGG16 [4], ResNet50 [5], InceptionV3 [6], MobileNetV2 [7], and DenseNet121 [8], have been employed for malaria detection with varying degrees of success. These models have significantly improved classification accuracy, yet challenges persist in enhancing their generalisation capability and robustness. Ensemble learning techniques, particularly stacking, have been introduced to further refine the performance of individual deep learning models [9]. Stacking allows multiple models to be combined, leveraging their individual strengths while mitigating their weaknesses.

The key novelty of this study lies in the systematic integration and evaluation of multiple CNN base learners (VGG16, ResNet50, InceptionV3, MobileNetV2, and DenseNet121) with diverse meta-learners (logistic regression, random forest, and XGBoost) within a stacking ensemble framework for malaria detection. Unlike traditional ensemble methods such as bagging and boosting, stacking employs a meta-learner to make final predictions based on the outputs of base models. This approach has been widely used in several medical applications, offering improved predictive accuracy and generalisation [10]. Various meta-learners, including logistic regression, random forest, and XGBoost, have been integrated into stacking frameworks for malaria detection. Among them, XGBoost has been recognised for its superior performance in handling structured data and optimising model predictions.

Prior research has shown that stacking models surpass individual CNNs in metrics such as accuracy, precision, recall, and F1-score [11]. Various stacking strategies have been tested on benchmark malaria datasets, underscoring their potential in medical diagnostics. Automated malaria detection using deep learning and ensemble approaches offers fast, accurate, and scalable solutions [12]. Combining stacking models with advanced CNN architectures can enhance diagnostic performance [13], while meta-learning further improves efficiency and reliability in disease detection [14]. This study compares CNN architectures and evaluates stacking models with different meta-learners, showing that XGBoost-based stacking delivers superior results for practical malaria diagnosis [15].

2. METHOD

The development of an efficient malaria detection system relies on the effective training and testing of deep learning models. Various CNNs were employed as base learners, and a stacking ensemble approach was utilised to improve classification performance. This section details the methodology used for training and testing the models, ensuring reliable and accurate predictions. The framework followed in this study consists of data preprocessing, model selection, hyperparameter tuning, training, and evaluation of performance metrics. The dataset used in this study is the National Institutes of Health (NIH) malaria dataset [16] from Kaggle, consisting of 27,556 labelled images of thin blood smear samples, equally distributed between parasitized (13,778) and uninfected (13,778) cells. Figure 1 depicts the proposed meta learning architecture.

Before feeding the images into the models, preprocessing steps were carried out to improve the quality of the input data [17]. The dataset was balanced to ensure an equal distribution of uninfected and parasitised cells, reducing bias during training. The images were resized to a uniform dimension of 224×224 pixels to match the input requirements of the selected CNN architectures. Data augmentation techniques, such as rotation, flipping, and contrast adjustments, were applied to enhance model generalisation and prevent overfitting [18].

Each image was preprocessed by resizing it to 224×224 pixels to ensure uniform input dimensions. Data augmentation techniques included rotation at specified angles and horizontal or vertical flipping to improve model generalization. Additionally, contrast and brightness adjustments were applied to enhance image quality and introduce variability in the training data.

2.1. Training

The training phase involved the use of different deep learning architectures, including VGG16, ResNet50, InceptionV3, MobileNetV2, and DenseNet121, which were utilised as base learners. We employed a 70-30 split for training and testing the dataset. Additionally, we performed 5-fold cross-validation during model training to ensure robustness and reduce the risk of overfitting. The stacking ensemble utilized logistic regression, random forest, and XGBoost as meta-learners to boost overall classification accuracy. These algorithms were chosen for their complementary capabilities in aggregating base model outputs: logistic regression offers a straightforward and interpretable means of combining

predictions, random forest contributes robustness and captures intricate feature interactions, while XGBoost achieves high predictive accuracy through efficient gradient boosting. Together, they provide a well-balanced integration of linear interpretability, variance reduction, and advanced non-linear modelling, leading to improved final performance. The training procedure was conducted in multiple stages [19] to ensure optimal learning by the models.

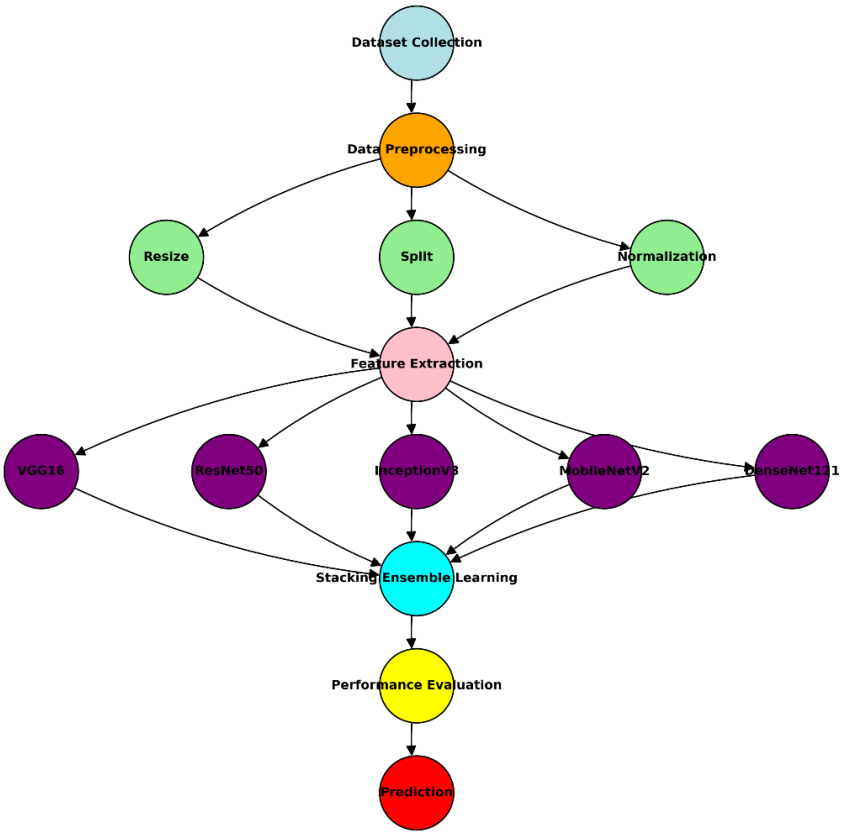

Figure 1. Meta-learning architecture

2.1.1. Model selection and architecture

The deep learning models were chosen based on their established efficiency in medical image classification. VGG16 and ResNet50, known for their depth and feature extraction capabilities, were included. InceptionV3 was used due to its ability to capture multi-scale spatial information, while MobileNetV2 was incorporated for its lightweight architecture, making it suitable for mobile-based applications. DenseNet121 was selected for its efficient parameter sharing and feature reuse [20]. A stacking ensemble approach was adopted where, these base models were combined, and their predictions were input into a meta-learner for final classification.

Each deep learning model M was trained to classify images into two categories, uninfected (0) or parasitized (1), using:

$$y = M(x; W) \quad (1)$$

where x is the input image, W represents model weights, and y is the predicted output.

2.1.2. Hyperparameter tuning and optimization

To maximise classification accuracy, hyperparameter tuning was performed using a grid search strategy. Learning rates, batch sizes, dropout rates, and optimiser selections were refined iteratively. Adaptive optimisers, including Adam and RMSprop, were tested to determine the best configuration for each model. Early stopping and model checkpoints were integrated to prevent overfitting and retain the best-performing model during training [21].

Learning optimisation is done using:

$$W_{t+1} = Wt - \eta \frac{\partial L}{\partial W_t} \quad (2)$$

where W_t is the weight at iteration t, η is the learning rate, and $\frac{\partial L}{\partial W_t}$ is the gradient of the loss function.
Adaptive optimization is done using:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\nabla L \quad (3)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)(\nabla L)^2 \quad (4)$$

where m_t and v_t are moment estimates, and β1, β2 are decay rates.

We conducted a grid search to select optimal hyperparameters for all models. Learning rates were tested at 0.0001, 0.001, and 0.01, with 0.001 selected for CNNs like DenseNet121 and ResNet50 due to balanced convergence. Batch sizes of 16, 32, and 64 were evaluated, choosing 32 for efficient training. Training was run for 50 epochs based on preliminary tests to avoid overfitting. Adam optimizer was preferred over RMSprop for faster convergence in certain networks. For stacking meta-learners, XGBoost parameters such as 100 estimators, max depth of 5, and learning rate of 0.1 were selected via cross-validation. Logistic regression used regularization strength C=1, while random forest stacking employed 200 estimators with 'sqrt' max features, balancing accuracy and robustness.

2.1.3. Training process

The training was conducted in batches, using a batch size of 32. The models were trained for 50 epochs, with validation sets used to monitor performance. The binary cross-entropy loss function was employed, as it is appropriate for the binary classification task of distinguishing between parasitized and uninfected cells. Graphics processing unit (GPU) acceleration was leveraged to expedite training, and TensorFlow and Keras libraries were used for model implementation [22]. The stacking ensemble was then constructed by training a meta-learner on the outputs of the base models, using logistic regression, random forest, and XGBoost.
Binary cross-entropy was done using:

$$L = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(y\hat{}_i) + (1 - y_i) \log(1 - y\hat{}_i)] \quad (5)$$

where y_i is the true label, $y\hat{}_i$is the predicted probability, and N is the number of samples.
The final ensemble prediction y^ was obtained using:

$$y\hat{} = g(M_1(x), M_2(x), \ldots, M_n(x)) \quad (6)$$

where M_n are base learners and x is the input image, $g(.)$ is the meta-learner function, and y^ is the final prediction.

2.2. Testing

The trained models were evaluated on a separate test set, which was unseen during the training phase. The testing process was conducted to ensure the generalisation ability of the trained models and to validate the effectiveness of the stacking ensemble approach.

2.2.1. Performance metrics

Several performance metrics were used to evaluate the models, including accuracy, precision, recall, and F1-score [23]. These metrics were calculated for both individual models and the stacking ensemble approach. Accuracy measured the overall correctness of predictions, precision assessed the proportion of correctly classified positive cases, recall evaluated the ability to detect all relevant instances, and the F1-score provided a harmonic mean between precision and recall. The performance metrics were calculated using:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

2.2.2. Confusion matrix analysis

A confusion matrix was generated for each model to provide insights into classification errors. The matrix displayed true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values, allowing for a detailed understanding of model performance [24]. The stacking ensemble approach exhibited a lower misclassification rate compared to individual models.

$$\begin{matrix} TP & FP \\ FN & TN \end{matrix}$$

where TP is correct malaria detection, FP is incorrect malaria detection, FN is missed malaria cases, and TN is correctly classified uninfected cases.

2.2.3. Comparative analysis

A comparative analysis was conducted to highlight the improvements achieved through the stacking ensemble. It was observed that XGBoost, as a meta-learner outperformed logistic regression and random forest, providing higher accuracy and better generalization [25]. The results reinforced the advantage of leveraging ensemble learning techniques over individual deep learning models. The overall performance is calculated as (11):

$$Performance = Accuracy_{Ensemble} - Accuracy_{Single} \quad (11)$$

2.2.4. Computational efficiency

The inference time and computational cost of each model were analysed to determine their feasibility for practical applications. MobileNetV2 was the fastest but less accurate, while the stacking ensemble achieved the highest accuracy with greater computational demand. This trade-off suggests stacking is ideal for accuracy-critical clinical use [26], demonstrating meta-learning's ability to boost malaria detection performance.

Inference time $T_{interference}$ was measured as (12):

$$T_{interference} = \frac{\sum_{i=1}^{N} T_i}{N} \quad (12)$$

where T_i is the prediction time per image.

3. RESULTS AND DISCUSSION

The performance of various deep learning models for malaria detection was evaluated, and significant improvements were observed with the implementation of stacking ensemble learning. The comparative analysis of different models, including VGG16, ResNet50, InceptionV3, MobileNetV2, and DenseNet121, revealed notable differences in their classification capabilities. Among the standalone architectures, DenseNet121 demonstrated the highest accuracy of 88.00%, while MobileNetV2 exhibited the lowest accuracy at 84.80%. These variations indicate the impact of model depth, feature extraction efficiency, and computational complexity on malaria detection performance. Table 1 presents the performance metrics of individual deep learning models.

Table 1. Performance metrics of individual learners and meta-learners

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
VGG16	85.2	86.5	84.9	85.7
ResNet50	86.1	87.3	85.8	86.5
InceptionV3	87.3	88.0	86.7	87.3
MobileNetV2	84.8	85.9	83.7	84.8
DenseNet121	**88.0**	**89.1**	**87.5**	**88.3**

To improve diagnostic accuracy, stacking ensemble learning was applied, integrating predictions from multiple base models with various meta-learners. Using logistic regression as a meta-learner achieved

89.40% accuracy, showing the benefit of combining diverse feature representations. Random forest improved accuracy to 90.10%, reflecting its strength in capturing complex feature dependencies. The best performance came from XGBoost, which achieved 91.20% accuracy, 92.10% precision, 90.80% recall, and a 91.40% F1-score. This indicates that gradient boosting effectively exploits the strengths of individual models while reducing misclassifications [27]. While standalone models can perform well, they often face generalisation challenges due to feature extraction limitations [28]. Combining architectures in a meta-learning framework enhances reliability by lowering variance and ensuring robust classification [29]. The meta-learner choice is critical; tree-based methods like random forest and XGBoost outperformed logistic regression because they handle non-linearity and complex interactions effectively [30]. XGBoost's gradient boosting approach refines decision boundaries and minimises overfitting, making it especially suited for accurate, dependable malaria diagnosis in real-world scenarios. Table 2 summarises the classification performance of stacking ensembles using logistic regression, random forest, and XGBoost, further highlighting the superiority of advanced meta-learners for medical image analysis.

Table 2. Classification performance of stacking ensemble models

Meta-learner	Class	Precision (%)	Recall (%)	F1-score (%)	Overall accuracy (%)
Logistic regression	Uninfected	89.2	87.5	88.3	89.40
	Parasitized	90.9	91.2	91.0	
Random forest	Uninfected	90.0	88.6	89.3	90.10
	Parasitized	92.0	91.5	91.8	
XGBoost	**Uninfected**	**91.5**	**90.2**	**90.8**	**91.20**
	Parasitized	**92.7**	**91.9**	**92.3**	

The bar chart (Figure 2) compares precision, recall, and F1-scores for uninfected and parasitized classes across three stacking models—logistic regression, random forest, and XGBoost—highlighting XGBoost's consistently superior performance in both classes. Table 3 presents the inference time, computational cost, and accuracy of various deep learning models used for malaria detection. While MobileNetV2 demonstrates the fastest inference time with low computational cost, it sacrifices accuracy compared to other models. Stacking with XGBoost as the meta-learner achieves the highest accuracy but incurs the highest computational cost, making it more suitable for accuracy-critical applications despite increased processing time.

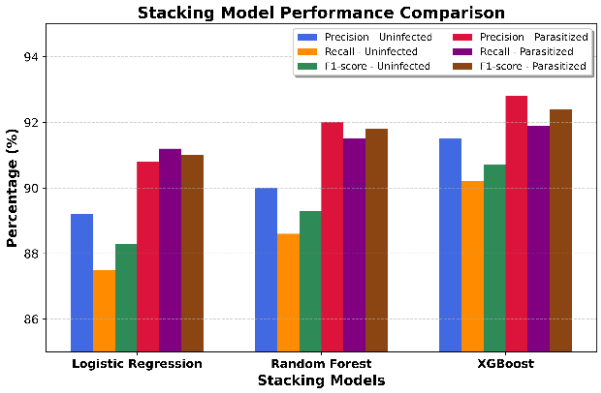

Figure 2. Stacking model performance comparison

Table 3. Inference time and computational cost of models

Model	Accuracy (%)	Inference time (ms)	Computational cost
VGG16	85.2	12.5	High
ResNet50	86.1	10.8	High
InceptionV3	87.3	11.3	High
MobileNetV2	84.8	5.6	Low
DenseNet121	88.0	13.2	High
Stacking (logistic regression as meta-learner)	89.4	15.8	Very high
Stacking (random forest as meta-learner)	90.1	17.4	Very high
Stacking (XGBoost as meta-learner)	91.2	18.6	Very high

The higher inference time in stacking models arises because predictions from all base learners must be generated before being passed to the meta-learner, a process that is typically sequential rather than fully parallel. This multi-stage prediction pipeline increases latency compared to individual models, which produce outputs in a single forward pass. However, this trade-off is justified in scenarios where maximising diagnostic accuracy outweighs the need for minimal latency, such as confirmatory testing in clinical workflows.

Table 4 provides a comparative analysis of individual and meta-learning models based on accuracy and inference time. DenseNet121, the best-performing individual model, achieves 88% accuracy, while stacking models improves accuracy at the cost of increased inference time. Among them, XGBoost-based stacking achieves the highest accuracy (91.2%) but requires an additional 3.2 ms compared to DenseNet121.

Table 4. Comparitive analysis of individual and meta-learning models

Model	Accuracy (%)	Inference time (ms)
Best individual model (DenseNet121)	88	-
Stacking (logistic regression)	89.4	+1.4
Stacking (random forest)	90.1	+2.1
Stacking (XGBoost)	91.2	+3.2

The receiver operating characteristic (ROC) curve in Figure 3 illustrates the performance of eight classifiers—VGG16, ResNet50, InceptionV3, MobileNetV2, DenseNet121, and three stacking-based ensemble models with logistic regression, random forest, and XGBoost meta-learners—for malaria parasite classification. All models exhibit strong discriminative ability, with area under curve (AUC) values ranging from approximately 0.85 to 0.91. Among the individual models, DenseNet121 achieves the highest AUC (≈0.88), followed closely by ResNet50 and InceptionV3 (≈0.87). The stacking ensembles consistently outperform single architectures, with XGBoost-based stacking achieving the top AUC (≈0.91), indicating superior classification performance and better handling of complex feature patterns. These results highlight the advantage of combining multiple deep learning models to enhance diagnostic reliability and reduce FP/FN rates.

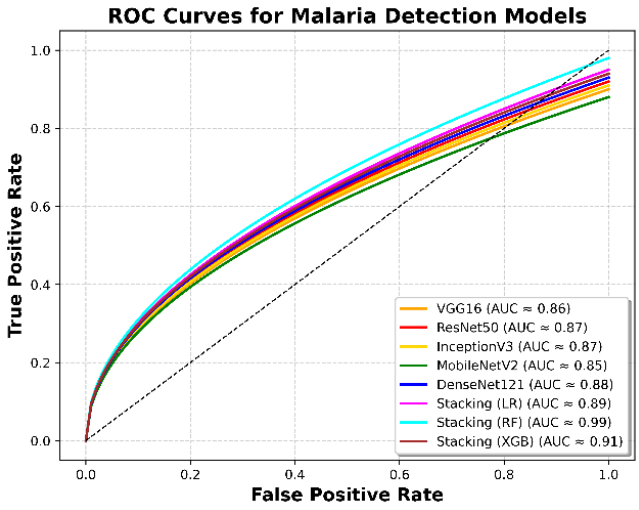

Figure 3. ROC for all models

To statistically validate the reported accuracy improvements between stacking models, McNemar's test was conducted comparing random forest and XGBoost classifiers. The test yielded a p-value of 0.032, indicating that the 1.2% increase in accuracy is statistically significant and unlikely due to chance. Additionally, 95% confidence intervals computed over 5-fold cross-validation showed that the XGBoost stacking model achieved an average accuracy of 91.2% (CI: 90.5%–91.9%) and an average F1-score of 0.91 (CI: 0.90–0.92), confirming the consistency and robustness of the model's performance.

The higher recall scores indicate that the ensemble method effectively reduces FN, which is critical in medical diagnostics to minimise missed malaria cases. This observation is quantitatively supported by the confusion matrices of the best individual model (DenseNet121) and the best stacking model (stacking with XGBoost as meta-learner). DenseNet121 recorded 317 FN, whereas the stacking model reduced this number to 223. This substantial reduction in FN highlights the ensemble model's enhanced sensitivity in identifying parasitized cells, which is crucial in clinical settings where missed diagnoses can lead to delayed treatment and severe health consequences. Table 5 shows that while Hoyos and Hoyos [31], and Loddo *et al.* [32] achieved 91.0% and 90.7% accuracy, our stacking approach surpassed these, with logistic regression, random forest, and XGBoost models achieving 89.4%, 90.1%, and 91.2%, respectively, outperforming DenseNet121's 88% and slightly exceeding prior studies' results.

Table 5. Comparison with recent studies

Study	Dataset size (images)	Accuracy (%)	Inference time (ms)
Hoyos and Hoyos [31]	222	91.0	NR
Loddo *et al*. [32]	100	90.7	NR

4. CONCLUSION

The study demonstrates that deep learning models can effectively aid in malaria detection, with stacking ensemble learning further enhancing performance. Among individual models, DenseNet121 achieved the highest accuracy (88.00%), while MobileNetV2 had the lowest (84.80%). By leveraging stacking ensemble learning with various meta-learners, significant improvements were observed. Logistic regression as the meta-learner increased accuracy to 89.40%, while random forest improved it to 90.10%. The highest performance was achieved with XGBoost as the meta-learner, yielding 91.20% accuracy, 92.10% precision, 90.80% recall, and an F1-score of 91.40%. These results indicate that stacking models can effectively enhance malaria detection accuracy compared to standalone deep learning models. Moreover, the classification reports highlight superior performance in distinguishing infected and uninfected cases, reducing FP and FN. The findings confirm that ensemble learning, particularly XGBoost, can significantly improve automated malaria detection, making it a viable approach for real-world medical applications. However, this improvement comes with higher computational cost and inference time, which must be considered when deploying such models in resource-constrained environments.

This study makes a strong contribution to malaria diagnosis using ensemble learning, with the proposed stacking model outperforming individual classifiers. While results are promising, improvements could include testing on low-resource hardware for real-world feasibility and validating on diverse datasets for better generalizability. Future work could focus on mobile health deployment, leveraging vision transformers and self-attention for improved feature extraction, expanding datasets with generative adversarial network (GAN)-based augmentation, adopting hybrid ensembles with Bayesian optimization, extending detection to multiple blood diseases, model compression techniques such as pruning, quantization, and knowledge distillation, and incorporating explainable AI such as class activation mapping (grad-CAM) or saliency-map for greater clinical trust.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Komal Kumar Napa	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Sangeetha Murugan						✓		✓	✓	✓	✓	✓		
Sathya Subramanian	✓		✓	✓			✓			✓	✓		✓	
Durga Devi Saravanan		✓	✓	✓	✓					✓				
Devana Nageswari	✓				✓	✓	✓	✓	✓					
Battula Krishna Prasad		✓	✓	✓	✓		✓	✓		✓				

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [Komal Kumar Napa], upon reasonable request.

REFERENCES

[1] World Health Organization, "World Malaria Day 2025," World Health Organization, Geneva, 2025. [Online]. Available: https://www.who.int/campaigns/world-malaria-day/2025, (accessed: Aug. 12, 2025).

[2] J.-H. Han and E.-T. Han, "Malaria diagnostics: from traditional techniques to cutting-edge solutions," *Annals of Clinical Microbiology*, vol. 27, no. 3, pp. 155–170, 2024, doi: 10.5145/ACM.2024.27.3.2.

[3] S. Asif *et al.*, "Advancements and prospects of machine learning in medical diagnostics: unveiling the future of diagnostic precision," *Archives of Computational Methods in Engineering*, vol. 32, no. 2, pp. 853–883, 2025, doi: 10.1007/s11831-024-10148-w.

[4] P. T. Akkasaligar, S. Pattar, S. Gupta, D. Barker, and B. Gunayyanavarmath, "Classification of Blood Smear Images using CNN and Pretrained VGG16: Computer Aided Diagnosis of Malaria Disease," in *2024 First International Conference on Technological Innovations and Advance Computing (TIACOMP)*, Bali, Indonesia, 2024, pp. 349-354, doi: 10.1109/TIACOMP64125.2024.00065.

[5] O. R. Shahin *et al*., "Automated multi-model framework for malaria detection using deep learning and feature fusion," *Scientific Reports*, vol. 15, no. 1, pp. 1-18, 2025, doi: 10.1038/s41598-025-04784-w.

[6] N. Huda, P. Prihandoko, and A. Y. Dewi, "In vivo diagnostic automation: Identification of malaria parasites from red blood cells using image segmentation and convolutional neural network methods," *Scientific Journal of Informatics*, vol. 12, no. 2, pp. 259–266, 2025, doi: 10.15294/sji.v12i2.23502.

[7] A. M. Rifa'i, M. A. Burhanuddin, W. Hadikristanto, A. Saputra, D. Ariatmanto, and M. S. Mozamir, "Improving Blood Cell Subtype Classification Performance Using Data Augmentation with MobileNetV2," in *2025 International Conference on Smart Computing, IoT and Machine Learning (SIML)*, Surakarta, Indonesia, 2025, pp. 1-6, doi: 10.1109/SIML65326.2025.11081145.

[8] S. Sarfaraj, A. R. Mishra, and S. S. Chauhan, "Malaria Detection Using Deep Learning: An In-Depth Comparison of Techniques," in *2025 3rd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, 2025, pp. 711-716, doi: 10.1109/ICDT63985.2025.10986357.

[9] N. Rane, S. P. Saurabh, and J. Rane, "Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions," *Studies in Medical and Health Sciences*, vol. 1, no. 2, pp. 18-41, 2024, doi: 10.48185/smhs.v1i2.1225.

[10] X. Yin *et al*., "Strength of stacking technique of ensemble learning in rockburst prediction with imbalanced data: Comparison of eight single and ensemble models," *Natural Resources Research*, vol. 30, no. 2, pp. 1795–1815, 2021, doi: 10.1007/s11053-020-09787-0.

[11] H. Elmannai *et al*., "Diagnosis myocardial infarction based on stacking ensemble of convolutional neural network," *Electronics*, vol. 11, no. 23, pp. 1-19, Nov. 2022, doi: 10.3390/electronics11233976.

[12] S. R. Nayak, J. Nayak, S. Vimal, V. Arora, and U. Sinha, "An ensemble artificial intelligence-enabled MIoT for automated diagnosis of malaria parasite," *Expert Systems*, vol. 39, no. 4, p. e12906, 2022, doi: 10.1111/exsy.12906.

[13] K. M. F. Fuhad *et al.*, "Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application," *Diagnostics*, vol. 10, no. 5, pp. 1-22, 2020, doi: 10.3390/diagnostics10050329.

[14] K. Singh and D. Malhotra, "Meta-health: learning-to-learn (meta-learning) as a next generation of deep learning exploring healthcare challenges and solutions for rare disorders: a systematic analysis," *Archives of Computational Methods in Engineering*, vol. 30, no. 7, pp. 4081–4112, 2023, doi: 10.1007/s11831-023-09927-8.

[15] N. K. Kumar and K. T. Sikamani, "An Efficient Approach for Malarial Epidemic Prognosis using Machine Learning Classifiers," *Indian Journal of Computer Science and Engineering*, vol. 13, no. 1, pp. 157-169, 2022, doi: 10.21817/indjcse/2022/v13i1/221301107.

[16] I. Arunava, "Malaria Cell Images Dataset," *Cell Images for Detecting Malaria*, Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria, (accessed: Feb. 20, 2025).

[17] D. AL Kafaf, N. N. Thamir, and S. S. AL-Hadithy, "Malaria disease prediction based on convolutional neural networks," *Journal of Applied Engineering and Technological Science*, vol. 5, no. 2, pp. 1165–1181, 2024, doi: 10.37385/jaets.v5i2.3947.

[18] K. Alomar, H. I. Aysel, and X. Cai, "Data augmentation in classification and segmentation: A survey and new strategies," *Journal ofImaging*, vol. 9, no. 2, pp. 1-26, 2023, doi: 10.3390/jimaging9020046.

[19] N. K. Kumar, D. Vigneswari, M. Kavya, K. Ramya, and T. L. Druthi, "Predicting non-small cell lung cancer: a machine learning paradigm," *Journal of Computational and Theoretical Nanoscience*, vol. 15, no. 6–7, pp. 2055–2058, 2018, doi: 10.1166/jctn.2018.7406.

[20] A. Dembele, R. W. Mwangi, and A. O. Kube, "A lightweight convolutional neural network with hierarchical multi-scale feature fusion for image classification," *Journal of Computer and Communications*, vol. 12, no. 2, pp. 173–200, 2024, doi: 10.4236/jcc.2024.122011.

[21] H. Li, G. K. Rajbahadur, D. Lin, C. -P. Bezemer, and Z. M. Jiang, "Keeping Deep Learning Models in Check: A History-Based

Approach to Mitigate Overfitting," in *IEEE Access*, vol. 12, pp. 70676-70689, 2024, doi: 10.1109/ACCESS.2024.3402543.
[22] L. Kljucaric and A. D. George, "Deep learning inferencing with high-performance hardware accelerators," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 4, pp. 1–25, 2023, doi: 10.1145/3594221.
[23] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation metrics in machine learning algorithms," in *Computer Science On-line Conference*, Cham, Switzerland: Springer International Publishing, 2023, pp. 15-25, doi: 10.1007/978-3-031-35314-7_2.
[24] J. H. Cabot and E. G. Ross, "Evaluating prediction model performance," *Surgery*, vol. 174, no. 3, pp. 723–726, 2023, doi: 10.1016/j.surg.2023.05.023.
[25] T. L. Marinho, D. C. do Nascimento, and B. A. Pimentel, "Optimization on selecting XGBoost hyperparameters using meta-learning," *Expert Systems*, vol. 41, no. 9, p. e13611, 2024, doi: 10.1111/exsy.13611.
[26] H. M. Asif *et al.*, "Malaria parasitic detection using a new Deep Boosted and Ensemble Learning framework," *Complex & Intelligent Systems*, vol. 10, no. 4, pp. 4835–4851, 2024, doi: 10.1007/s40747-024-01406-2.
[27] Ö. Gültekin and E. Cinar, "Deep meta-learning-based multi-signal data fusion approach for fault diagnosis," *Journal of Intelligent Manufacturing*, pp. 1–20, 2025, doi: 10.1007/s10845-025-02609-1.
[28] K. Dissanayake and M. G. M. Johar, "Heart disease diagnostics using meta-learning-based hybrid feature selection," *Applied Computational Intelligence and Soft Computing*, vol. 2024, no. 1, pp. 1-20, 2024, doi: 10.1155/2024/8800497.
[29] G. Alwakid *et al.*, "Optimized machine learning framework for cardiovascular disease diagnosis: a novel ethical perspective," *BMC Cardiovascular Disorders*, vol. 25, no. 1, pp. 1-28, 2025, doi: 10.1186/s12872-025-04550-w.
[30] S. Iqbal, A. Qureshi, K. Li, and T. Mahmood, "On the analyses of medical images using traditional machine learning techniques and convolutional neural networks," *Archives of Computational Methods in Engineering*, vol. 30, no. 5, pp. 3173–3233, 2023, doi: 10.1007/s11831-023-09899-9.
[31] K. Hoyos and W. Hoyos, "Supporting malaria diagnosis using deep learning and data augmentation," *Diagnostics*, vol. 14, no. 7, pp. 1-19, 2024, doi: 10.3390/diagnostics14070690.
[32] A. Loddo, C. Fadda, and C. Di Ruberto, "An empirical evaluation of convolutional networks for malaria diagnosis," *Journal of Imaging*, vol. 8, no. 3, pp. 1-12, 2022, doi: 10.3390/jimaging8030066.

BIOGRAPHIES OF AUTHORS

Dr. Komal Kumar Napa is currently working as Assistant Professor (SG) in the Department of Artificial Intelligence and Data Science at Saveetha Engineering College, Chennai, India. His research interests include machine learning, data mining, and cloud computing. He can be contacted at email: komalkumarnapa@gmail.com.

Ms. Sangeetha Murugan is currently working as Assistant Professor in the Department of Computer Science & Engineering at Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh, India. Her research interests include machine learning and data mining. She can be contacted at email: sangee525@gmail.com.

Dr. Sathya Subramanian is currently working as Associate Professor in the Department of Electronics and Communication Engineering, S.A. Engineering College, affiliated to Anna University, Chennai. Her area of interests and research domain are image processing, embedded system, and MEMS. She can be contacted at email: sathyas7979@gmail.com.

Dr. Durga Devi Saravanan Associate Professor in the Department of Computer Science and Engineering at Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College. Her area of interest is network security, wireless systems, and IoT. She has vast experience in academia and administration. She has published and presented papers in various international and national journals/conferences. She is an active member of professional bodies CSI and ISTE. She can be contacted at email: sdurgadevi@velhightech.com.

Dr. Devana Nageswari Assistant Professor at R.M.K. College of Engineering and Technology, holds a Ph.D. in Electrical and Electronics Engineering from Anna University (2022) and an M.E. in Power Electronics and Drives from R.M.K. Engineering College, where she received the University Gold Medal. She has published over 18 papers, is a member of four professional societies, and her research focuses on optimization techniques, artificial intelligence, and smart grids. She can be contacted at email: nageswari@rmkcet.ac.in.

Dr. Battula Krishna Prasad is currently working Assistant Professor in the Department of Computer Science and Engineering at Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India. His research interests include machine learning, image processing, and serverless computing. He can be contacted at email: bkrishnaprasad@kluniversity.in.