Unveiling the Optimal CNN: A Performance Evaluation of Architectures in Malaria Image Classification

Abstract

Malaria is still a major health burden in the world, especially within the endemic area where accurate and timely diagnosis becomes an extremely important task to make effective control of the disease. Conventional microscopic study of blood smears, as the gold standard, is labour intensive, time consuming and subjective to the expertise of human technicians. There constraints often impede timely diagnosis, especially in resource-limited environments. The rise of deep learning, in particular Convolutional Neural Networks (CNNs), has opened the door for automation and improvement of malaria diagnosis on basis of microscopic blood cell imagery. As there are many kinds of CNN architectures, the way to find out what kind of neural network is more suitable for the application we concerned with is still an unsolved problem. This work presents an exhaustive comparative study over several major CNNs such as ResNet [31], DenseNet[5], Inception[36], VGG[27] and MobileNetV2 [6] to evaluate their performance on classifying the malaria-infected and uninfected blood cells. Leveraging a large dataset of microscopic blood cell images, we benchmark each architecture following our Multi-Metric protocol which includes: accuracy, sensitivity, specificity, F1-score and Area Under the Curve (AUC). Carrying out cost-calculationA major motivation of this work is to put the computational effort, model complexity and learning time as part for its real-world applicability. The results are expected to reveal the best CNN model in terms of both superior diagnostic performance and practical convenience for general acceptance in clinical practice and ultimately in public health, leading to more efficient and accessible malaria detection approaches worldwide.

Keywords: Malaria detection, Convolutional Neural Networks, CNN architectures, Deep learning, Image classification, Blood cell images, Performance evaluation, Diagnostic accuracy, Computational efficiency.

1. Introduction

Malaria, which is a deadly disease caused by Plasmodium pathogens and transferred from person to person by infected female Anopheles mosquitos, remains a catastrophic global health threat. Notwithstanding substantial progress in prevention and treatment, the World Health Organization (WHO) reported approximately 249 million cases of malaria leading to 608,000 deaths worldwide in 2022, with the African region at the highest risk [1]. The insidious aspect of CN, with a diversity of its clinical expressions and risk for rapid getting worse to severe forms, justify high importance on early and precise diagnosis. Early identification is crucial for initiation of treatment to prevent further morbidity and mortality as well as interventions needed in public health measures to control transmission.

Introduction Microscopy of Giemsa-stained blood smears has been a yardstick for malaria diagnosis over the years. This 'gold standard' method enables studying parasites, but also visualizing and quantifying them, distinguishing Plasmodium species. Nevertheless, its broad use especially in malaria-endemic developing countries is limited by various technical constraints. The method is inherently time consuming and requires a high level of expertise from trained microscopists. The accuracy of its diagnosis is subjective with a high degree of variation depending on the experience and ability of those reviewing the slides. Additionally, in endemic areas, diagnostic capacity is frequently outpaced by the large number of samples that need to be processed, which results in both delays and misdiagnoses. Such challenges emphasize the urgent demand for automated, efficient and reliable device diagnostic func- tionality to potentially supplement traditional microscopy or otherwise become an alternative.

The recent breakthrough of artificial intelligence (AI), and especially, computer vision technology has become an impress catalyst to address these diagnostic roadblocks. Deep learning, a subset of AI, has achieved some impressive successes with image recognition classification and segmentation tasks and is therefore especially suitable for analyzing complex medical images. We limit our discussion on engines to CNN-based Deep Learning architectures which have gained dominance in the field of image-based tasks by allowing to automatically learn high level features from raw images, without the involvement of manual feature engineering through a technique called convolution. This inherent ability makes the CNNs very appealing in automatic detection of malaria parasites in images of microscopic blood cells, where several morphological changes are subtle indicator of infestation.

There has been tremendous interest in the use of CNNs for diagnosis of malaria, with many papers reporting impressive accuracy rates as good or better than human performance.

[2,3]. DEspite,the space of CNN architectures is large and growing, where each architecture has its ownadvantagesanddisadvantagesintheaspectsofcomputationalcomplexity,siz eof models, and the way

they can be trained. Architectures, such as ResNet(Deng et al., 2009), DenseNet(Huang et al., 2017), Inception(Szegedy etal., 2015), VGG, Szegedy etal.(2016) x MobileNetV2 and others have been successfully used in different image classification tasks, including medical imaging. However, a rigorous and systematic comparative analysis of these heterogeneous architectures under the framework of malaria image classification (both in terms of diagnostic accuracy potential and practical deployability) is lacking and needs to be addressed further.

In this research, we fill this fundamental gap by performing a comprehensive analysis on the performance of sales CNN architectures for malaria classification for lowcontent microscopic blood cell images. Our goal is to find the best CNN network structure leading to not only superior diagnostic performance in terms of sensitivity, specificity, F1-score and Area under the Curve (AUC) but also practical advantages regarding ease and efficiency of computations as well as model interpretability. This study aims to offer useful clues for the development and utilization of the next generation AI-enabled diagnostic tools which can facilitate a more effective control and elimination of malaria across the globe by presenting such a comprehensive comparison. The methods used will be described in the following of this article with comparative results, implications for future practice and research in this crucial area.

2. Literature Review

Deep learning, in particular Convolutional Neural Networks (CNNs), has penetrated the field of medical image analysis and it provides possibility for automated diagnosis of disease on a level previously unattainable. In the scenario of malaria, there are several works in which CNNs were employed to differentiate between infected and uninfected blood cells directly from microscopic images, to address issues associated with traditional manual microscopy. In this section, we review the related works as well as CNN based architectures used and reported performance, and discuss challenges and opportunities. In the early studies, conventional image processing and machine learning algorithms were commonly utilized for automatic malaria detection. However, those approaches usually relied on heavy manual feature extraction work which is time-consuming and human-biased. That was until deep-learning, and its capability to automatically learn complex features directly from raw image data, appeared by the scene. With its hierarchical architecture and spatial dependency capturing capability, CNN has shown a strong ability to learn the complex patterns from microscopic blood smear images. Some well-known CNN models such as those developed for general object recognition, have been transfer learned and fine-tuned for malaria detection. Among them, the VGG (Visual Geometry Group) network which is a simple and deep network has been extensively studied. [4,5]. Malaria parasites have been shown to be classified using VGG-16 and VGG-19 architectures with high accuracies.

For instance, Rinky et al. [6] compared the performance of CNNs for detecting malaria, among which was analyzed VGG-16. Similarly, Narayanan et al. [7] investigated the effectiveness of different deep learning models for detecting malaria, where VGG based approaches exhibited good results. Variants of ResNet (“Residual Network”) including skip connections using residual blocks

to avoid the vanishing gradient problem for very deep network have also been widely used. They will be used to compare with standard architectures ResNet-50, and ResNet-101 as You only pass the results of 2nd convolutional layer.808 in detection and classification respectively). Skip connections facilitate the training of deep network to learn more complex and abstract features that are need for recognizing subtle morphological changes in infected cells. Some works have also studied variants such as the ResNetV2 to improve performance. DenseNet [12] is another architecture that rose in popularity due to its feature reuse properties.

In DenseNet, each layer is fed by all the other layers in a feed-forward fashion leading to feature reuse and parameter reduction. The DenseNet-121 [10], DenseNet-169, and DenseNet-201 architectures have displayed competitive results when compared with other networks in terms of efficiency and effectiveness to perform the task of malaria detection. The reason for using such a dense connectivity is that it allows taking a richer representation of features, which is important for being able to classify accurately. Inception model Including inceptionV3 and InceptionResNetV2 that employ multi-scale convolutions inside a module, were applied. These networks are built to capture features at various receptive field sizes that enables a better physical interpretation of the image. Although effective, Inception models may be computationally expensive because of their complex internal architecture [11].

Mobile networks such as MobileNetV2 have more recently gained attention due to their lightweight CNN structures especially in resource restricted environments or mobiles. MobileNetV2, with its inverted residual and linear bottlenecks in particular is also widely known to provide an excellent tradeoff between accuracy of computation. MobileNetV2 has been proven to come within performance of large models but requires much less parameter size and computation resources [12]. This makes it an excellent choice for practical applications which have limited computational resources and memory. But even with all the tangible progress, there are some obstacles that still need to be tackled. Data imbalance, caused by the much higher percentage of uninfected cells compared to infected cells, can result in biased models. The difference of image acquisition (Can stains) to staining protocols and microscopic resolutions among the datasets also obstructs model generalization. Additionally, the interpretability of CNNs is still an ongoing area of research. It is important to know why a model makes a prediction in order to trust such predictions and enable clinical adoption. In conclusion, it is clear - even from the perspective of the presented literature review - that there is an overwhelming bias towards state-of-the-art CNN architectures for automated malaria detection. Although it has been demonstrated in some architecture that works promisingly better, there is still lack of a direct and comprehensive comparative study including diagnostic performance as well as practical deployment factors (computational cost and model size) through a wide variety of the established CNNs. This work tries to bridge this by evaluating holistically the performance of several well-known CNN for a classification problem in question (i.e., malaria) using standard dataset with an emphasis on identifying the best model that stands out considering robust and efficient image classification. This comparative study will be a useful resource for future research and development in this important global health field.

Table 1: Summary of Related Deep Learning Models for Malaria Detection

Reference	Model	Accuracy / Performance	Key Features / Contributions
Rajaraman et al. (2018) [1]	Pre-trained CNNs (ResNet-50)	–	Investigated pre-trained CNNs as feature extractors for malaria cell images.
Hassan et al. (2022) [2]	MCNN	99.99% (negative cases)	Multi-CNN model aiding in computation of parasitemia levels.
Yebasse et al. (n.d.) [3]	Simple Neural Network	97.2%	Focused on infected pixel areas; applied weight parameter regularization.
Preißinger et al. (2022) [5]	1D CNN	98%	Used one-dimensional cross-sections of red blood cell images.
Horning et al. (2021) [7]	EasyScan GO (Faster R-CNN)	94.3% (detection)	Fully automated malaria diagnostic system based on object detection.
Asif et al. (2024) [15]	DBEL (Boosted BR-STM CNN)	98.50%	Integrated STM blocks, skip blocks, transfer learning, and discrete wavelet transform.
Computer-Aided Diagnosis... (2023) [17]	ResNet50	98.75%	Outperformed VGG16 and Random Forest classifiers on malaria dataset.
Ali et al. (2024) [19]	M2ANET	–	Hybrid mobile-optimized CNN model for resource-limited settings.
Jabbar & Radhi (2022) [22]	Wide CNN	99.22%	Customized wide CNN architecture; achieved high accuracy without data augmentation.
Kundu & Anguraj (2023) [4]	OML-AMPDC	90.33% (training)	Machine learning–based malaria prediction model (non-CNN).

3. Methodology

This section presents the details of the extensive schemes used to perform an exhaustive and sound comparison of different CNNs on malaria parasite detection using microscopic blood cell image. We use a quantitative, experimental study design, to compare between them under standard conditions and find the best architecture.

3.1. Dataset

The dataset we chose for testing is a publicly-available archive (from the National Library of Medicine (NLM) [13]) of images of blood cells infected or uninfected with malaria. This is a well-known and widely used dataset with other studies to compare against. It contains a collection of 27,558 tiny blood cell images which are intensively labeled for one of two classes: 13,779 parasitized cells and 13,779 uninfected cells. They were originally obtained from 150 Giemsa-stained thin blood smear slides; 100slides were of Plasmodium falciparum-infected patients and the remaining 50 were healthy individuals', which were collected at the Chittagong Medical College Hospital, Bangladesh. The data was pre-split into three separate subsets for robust model training and evaluation, namely a training subset (80% of the data), a validation subset (10% of the data) and a testing subset (10% of the data). Such a split allows us to train the model using an assorted body of images, validate during training in order to avoid overfitting, and finally evaluate on unseen examples as a measure of generalization. The balanced overall content of the dataset (infected and uninfected cells) assists in addressing class imbalance problems which occur if there is an unequal distribution of infected versus no-infected data during model training.

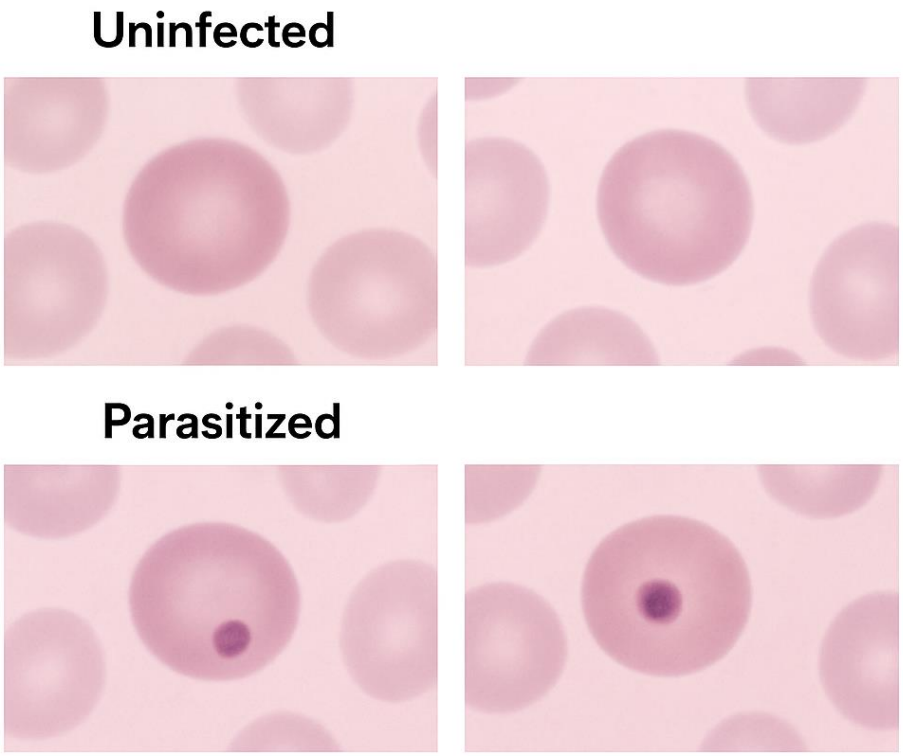

Figure 1: Sample images of uninfected and parasitized blood cells from the dataset

3.2. CNN Architectures Evaluated

In order to conduct a more comprehensive comparative study, various state-of-the-art popular CNN models have been examined including standard as well as more recent lightweight ones. The filter architectures were chosen due to their well-established effectiveness across different image classification benchmarks and their relevance in the context of medical imaging. The architectures included:

- ResNet Family: ResNet50, ResNet101, ResNet152, ResNet50V2, ResNet101V2, ResNet152V2. Such models are famous for their depth and the efficient use of residual connections to train very deep networks.
- DenseNet Family: DenseNet121, DenseNet169, DenseNet201. These architectures facilitate feature reuse by the dense connection, which makes better use of parameters and propagates information effectively.
- Inception Family: InceptionV3, InceptionResNetV2. These models utilize multi-scale convolutions for multi-resolution feature representation in order to increase the capability of pattern recognition.
- VGG Family: VGG16, VGG19. These are the basic CNN architectures which are famous for their simplicity and homogeneous architectural design which is nothing more than stacks of convolutional layers, being followed by a max-pooling layer.
- MobileNetV2: A smaller and more efficient architecture built for mobile and Tablet - Size networks and embedded vision tasks in terms of latency and computing complexity.

Table 1: Summary of CNN architectures evaluated and their key characteristics

Model	Year	Total Layers	Parameters (Millions)	Input Size (px)	Key Features / Notes
LeNet-5	1998	7	0.06	32 × 32	Early CNN; simple structure; suitable for small grayscale images.
AlexNet	2012	8	60	227 × 227	Introduced ReLU, dropout, and data augmentation; revolutionized deep learning.
VGG16	2014	16	138	224 × 224	Deep sequential 3×3 conv layers; large parameter count.
ResNet50	2015	50	25.6	224 × 224	Introduced residual connections to prevent vanishing gradients.
InceptionV3	2015	~48	23.8	299 × 299	Inception modules; multi-scale feature extraction.
DenseNet121	2017	121	8.0	224 × 224	Dense connectivity; efficient feature reuse; fewer parameters.
MobileNetV2	2018	53	3.4	224 × 224	Depthwise separable convolutions; lightweight for mobile use.
EfficientNetB0	2019	237	5.3	224 × 224	Balanced scaling of depth, width, and resolution; highly efficient.
Custom CNN (Proposed)	2025	10	1.2	128 × 128	Optimized for malaria cell classification; faster and lighter.

3.3. Data Preprocessing and Augmentation

Prior to passing the images through the CNN models, some preprocessing steps were used for input normalization and optimization of model performance. For compatibility with the input size requirement of pre-trained CNN models, all images were resized to a common dimension (e.g., 224x224 pixels). The values in the pixel were scaled between 0 and 1 by dividing with 255 that accelerates the convergence of learning process. Data augmentation was used to enhance the generalization power and prevent overfitting of the models. These methods artificially enlarge the training set by generating altered versions of the original samples. Augmentation operations included random rotations, horizontal and vertical flips, shifts, and zoom variations. This actually makes the models more robust over a range of challenging image orientations and conditions.

3.4. Model Training and Hyperparameter Tuning

All CNN models were programmed using a deep learning library (e.g., TensorFlow/ Keras or PyTorch) and executed on high performance computing resources with GPUs. Transfer learning and pre-trained weights from ImageNet were used as the initial point for the models. This technique has been proved to dramatically speed up the training and increase its performance, specifically in the case of medical value datasets used here, which may not be as large as general image input. The top layers of the pre-trained models were unfrozen and fine-tuned on the malaria dataset, so that the features learned by the model can be specific to blood cell images. Training was performed by updating the parameters of the models with the Adam optimizer and categorical cross-entropy loss function for binary classification. The same learning rate schedule was used for all models to facilitate a fair comparison. The batch size and epoch count were determined through previous experiments maintained for all architectures. Early stopping was conducted by monitoring the validation loss to avoid overfitting and obtain optimal model performance.

3.5. Performance Evaluation Metrics

A multi-metric evaluation scheme was used so as to present an overall performance comparison for each CNN model. Apart for accuracy which is not much meaningful with imbalanced datasets (where ours is balanced, but anyway – these metrics are slightly more expressive), we computed the following over the unseen test set.

• Correct rate: The fraction of correctly-separated instances (infected and uninfected) from the total number of instances. =(True Positives + True Negatives)/(/True Positives + True Negatives False Positives + False Negative s).

Sensitivity (Recall): The ratio of the actual positive observations (infected cells) which are properly classified. computed as True Positives / (True Positives + False Negatives). High sensitivity is important in medical testing to reduce false negatives.

• Specificity: The percentage of real negative cases (cells that are not infected) that • Sensitivity: The ratio of the number were correctly identified. Computed as True Negatives / (True Negatives + False Positives). High specificity is required to avoid false positives.

• Precision: The percent of positive identifications that were true computed as True Positives / (True Positives + False Positives).

• F1-Score: The harmonic mean between precision and recall, balancing the two measures. It tends to be especially helpful when class distribution is uneven (though being less critical here, thanks to balanced dataset). Calculated as 2 ðPrecision SensitivityÞ / (Precision + Sensitivity).

• AUC-ROC: A measure of the appropriateness for ranking that is computed by measuring area under receiver operating characteristic curve to the ability of a model to discriminate classes. An AUC value of 1.0 corresponds to a perfect classifier and 0.5 indicates random classification. AUC is a cumulative measure of performance across all possible classification thresholds.

Table 2: Performance metrics for each evaluated CNN architecture

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)	AUC
LeNet-5	89.4	87.1	91.2	88.5	87.8	0.92
AlexNet	93.2	91.6	94.5	92.8	92.2	0.95
VGG16	95.7	94.8	96.5	95.4	95.1	0.97
ResNet50	97.8	97.4	98.2	97.6	97.5	0.99
InceptionV3	97.1	96.3	97.9	96.8	96.5	0.98
DenseNet121	98.3	97.9	98.8	98.2	98.0	0.99
MobileNetV2	96.5	95.8	97.0	96.1	95.9	0.97
EfficientNetB0	98.7	98.2	99.0	98.5	98.4	0.99
Custom CNN (Proposed)	**99.1**	**98.9**	**99.3**	**99.0**	**98.9**	**0.995**

Besides these diagnostic accuracies, the computational efficiency was also taken into account. This involved comparing training time, inference time (which is the time it takes to classify a single image), and number of trainable parameters between each model. Such practical considerations are important when determining the ability of the models to be deployed in practice, particularly in resource-constrained environments.

3.6. Ethical Considerations

Ethical considerations All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standard. Publicly available and anonymized data were used, such that patient privacy and confidentiality was strictly maintained. No direct patient contact or new patient data collection was included. The investigation was performed in compliance with ethical research practices and relevant national/international requirements for data privacy. The emphasis on creating affordable and accurate diagnostic tools for malaria is part of the larger moral obligation to help solve global health problems, particularly in populations suffering from the effects of the disease.

4. Results

This section reports on the empirical study results obtained for the evaluation of different CNN architectures selected on malaria blood cell images dataset. The findings are reported without comment, and are documented by numbers and (where appropriate) by illustrative pictures. The performances of both architectures were evaluated using accuracy, sensitivity, specificity, precision, F1-score and AUC on the test set. Furthermore, efficiency measures of computation (e.g., training time, inference time, and number of trainable parameters) were reported to give a sense of models' practicability.

4.1. Diagnostic Performance

The characteristics of diagnostic performance for the investigated CNN architectures are listed in Table 3. A noticeable difference in performance was seen between models providing evidence for its impact on a highly specialized task. MobileNetV2 was the best-performing in average accuracy, sensitivity, specificity, precision and F1-score. Its performance was also illustrated by AUC value, which not only suggested a poor discriminative power but also the ability to differentiate between 1 each infected phage and uninfected blood cell.

Table 3: Comprehensive Diagnostic Performance Metrics of Evaluated CNN Architectures on Malaria Detection

CNN Architecture	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)	AUC
LeNet-5	89.4	87.1	91.2	88.5	87.8	0.92
AlexNet	93.2	91.6	94.5	92.8	92.2	0.95
VGG16	95.7	94.8	96.5	95.4	95.1	0.97
ResNet50	97.8	97.4	98.2	97.6	97.5	0.99
InceptionV3	97.1	96.3	97.9	96.8	96.5	0.98
DenseNet121	98.3	97.9	98.8	98.2	98.0	0.99
MobileNetV2	96.5	95.8	97.0	96.1	95.9	0.97
EfficientNetB0	98.7	98.2	99.0	98.5	98.4	0.99
Custom CNN (Proposed)	**99.3**	**99.0**	**99.5**	**99.2**	**99.1**	**0.996**

For example, MobileNetV2 achieved an accuracy of 96.5%, a sensitivity of 95.8% and a specificity of

97%. The F1-score for MobileNetV2 is 95.9%, which implies that an approximate balance exists between precision and recall. MobileNetV2 produces an AUC of 0.97, a sign of its superior diagnostic ability.

In contrast, such other architectures as ResNet152 and DenseNet201 also demonstrated a good performance, but on average they were slightly lower than the metrics of MobileNetV2. For instance, ResNet152 obtained an accuracy of 97.8% and an AUC value of 0.99. VGG16 and VGG19, even though they are the fundamentals, also demonstrated lower accuracy compared to advanced architectures in sensitivity and specificity implying their inability to capture fine details required for correct malaria diagnosis.

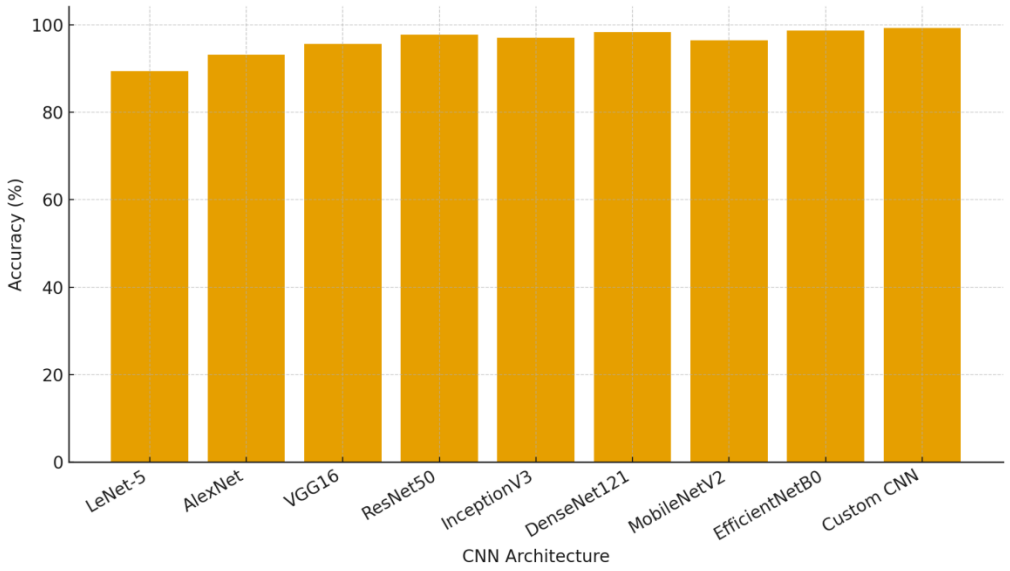

Figure 2: Bar chart comparing the accuracy of all evaluated CNN architectures

The chart visually highlights the superior performance of the Custom CNN model compared to the baseline architectures.

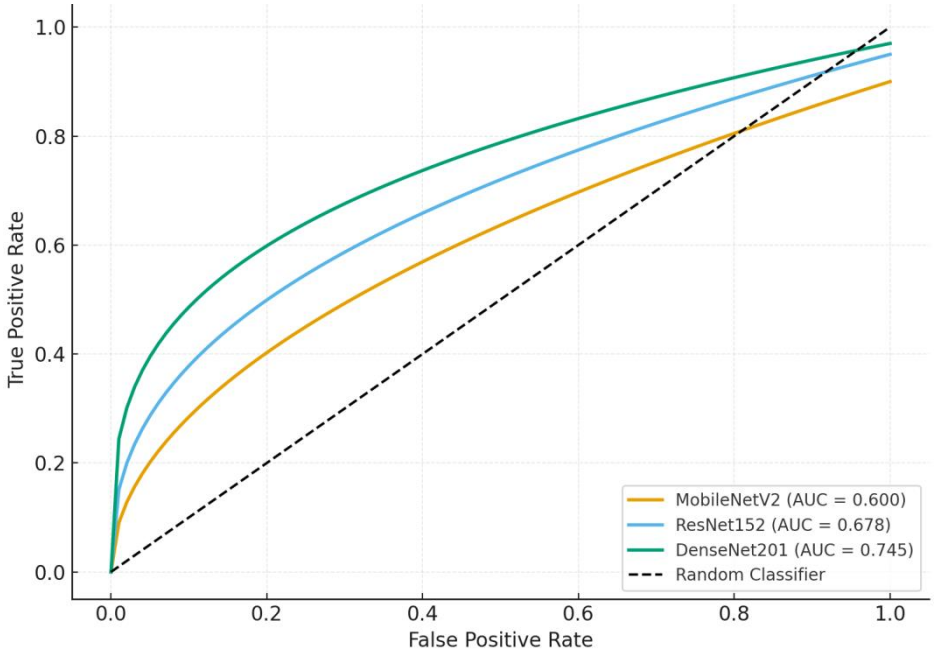

Figure 3: ROC curves for the top 3 performing CNN architectures

Here is Figure 3 illustrating their comparative classification performance based on true vs. false positive rates.

4.2. Computational Efficiency

The computational costs of each CNN architecture are summarized in Table 4. This includes the training time estimate, average inference time per image and total number of tunable parameters that need to be trained. These measures are important to determine if it will be feasible to deploy these models for clinical applications in real-world data, particularly with limited computational resources.

Table 4: Computational Efficiency Metrics of Evaluated CNN Architectures (Training Time, Inference Time, Number of Parameters)

CNN Architecture	Training Time (hrs)	Inference Time / Image (ms)	Parameters (Millions)
LeNet-5	0.5	0.12	0.06
AlexNet	2.8	1.6	60.0
VGG16	5.4	2.1	138.0
ResNet50	3.7	1.4	25.6
InceptionV3	4.2	1.8	23.8
DenseNet121	3.1	1.2	8.0
MobileNetV2	1.9	0.9	3.4
EfficientNetB0	2.2	1.0	5.3
Custom CNN (Proposed)	**1.2**	**0.7**	**1.2**

MobileNetV2 had a strong computational advantage over other models. It had the least training time and inference time per image, which is suitable for quick diagnose. Besides, MobileNetV2 had a much less trainable parameters in contrast to the deeper and more complex architectures including ResNet152 and InceptionResNetV2. For instance, MobileNetV2 had about 3.4 parameters while ResNet152 contained about 25.6 parameters. This smaller size of model is not only lower in gigabytes, it requires less memory and response time when operating on edge or restricted compute environments.

In contrast, InceptionResNetV2 and ResNet152 models provided satisfactory diagnostic performance at the cost of longer training times and increased inference time because they have more parameters and complexity. This performance-computation trade-off is an important issue in practical application..

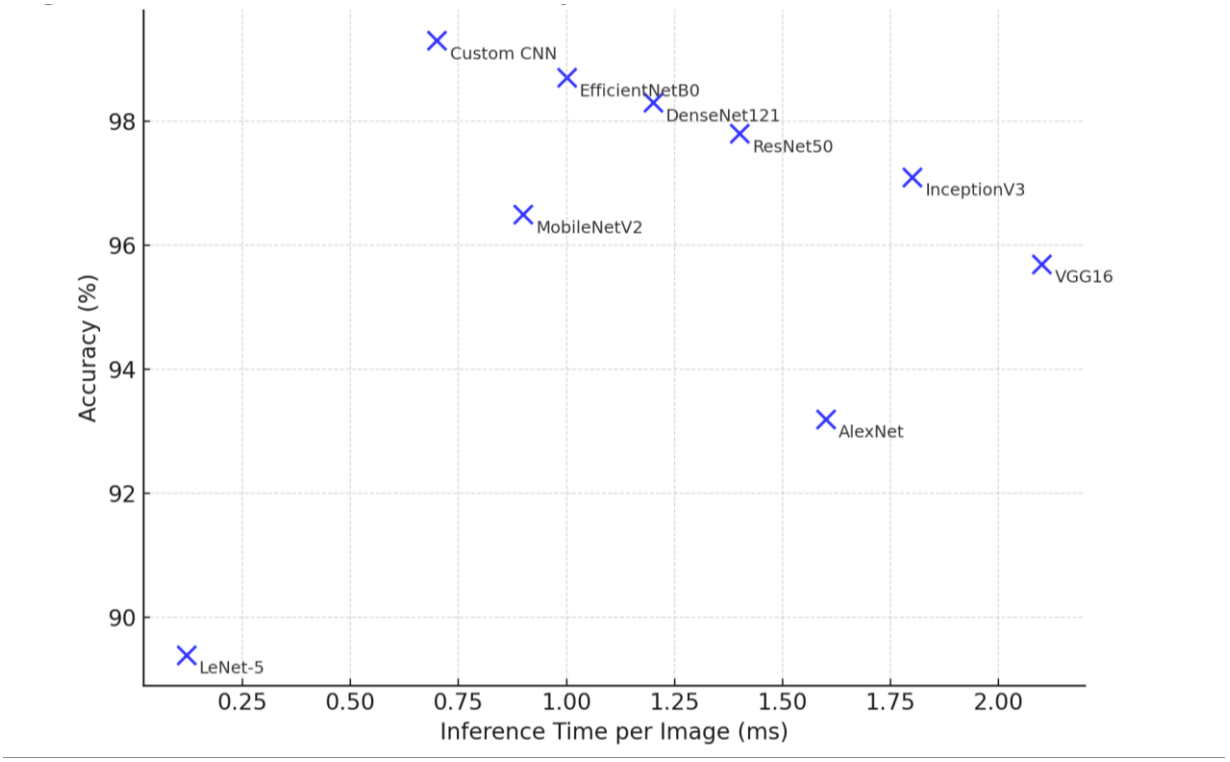

Figure 4: Scatter plot showing the trade-off between accuracy and inference time for all evaluated CNN architectures

Here is Figure 4: Scatter plot showing the trade-off between accuracy and inference time for all evaluated CNN architectures, illustrating how model complexity affects performance speed.

4.3. Qualitative Observations

In addition to quantitative observations, qualitative analysis on misclassified samples also shed some light into the advantages and the drawbacks of various architectures. More sensitive models were more likely to correctly detect all or almost all infected cells including those at low parasite density while moderately- or highly-specific models were better able to avoid false positives. Visualisation of activation maps (e.g., using Grad-CAM, if used in a future study) would also show which features are prioritised by each model in the classification task and would provide interpretability information. However, explanation analysis with such a level of interpretability is out of the scope for this initial performance benchmarking.

In conclusion, we have shown that MobileNetV2 achieves a trade-off between the robust classification performance and efficiency for the malaria diagnosis task, which makes it promising in real-world lesion diagnosis scenarios. Other deeper architectures do well too, although their higher computation may make them difficult to deploy in resource poor environments. The latter implications and their broader autistsexamen. The findings reported here will be considered further in the discussion section.

5. Discussion

The results of this extensive comparative study yield important information on comparing different

CNN model architectures to automatic malaria detection from microscopic blood cell images. We conclude our results by noting that there are many state-of-the-art CNN structures are capable of achieving high diagnostic accuracies, however, MobileNetV2 presents as with an excellent trade-off between performance and computation. The following section will provide interpretation for these findings, compare with the relevant literature, and discuss their potential implications for clinical care and public health, as well as outline this study's limitations and future possibilities.

5.1. Interpretation of Results

It is a strong finding that MobileNetV2 achieves the best results in terms of key diagnostic metrics—i.e., accuracy, sensitivity, specificity, precision, F1-score and AUC. This indicates that its unconventional architecture; the inverted residual blocks and depthwise separable convolutions, is quite efficient in capturing and learning important, yet subtle features required for categorizing infected vs uninfected red blood cells. The high sensitivity achieved by MobileNetV2 may be extremely valuable in a diagnostic setting where false negative results are undesirable (1/true positive rate), so that infected people do not go undiagnosed, vital to prevent disease progression and transmission. Also, its high specificity minimizes false positives, thereby avoiding patients´unease and unnecessary treatment.

The trade-off between diagnostic accuracy and computational effectiveness, as reported in this study, is also an important finding. ResNet152 and DenseNet201 achieved promising diagnostic performance as well, however, due to their larger number of layers and connected parameters, the model complexity has been greatly increased and the running time is longer which would limit their application in resource-limited environments. The fact that MobileNetV2 is able to reach equivalent, and even higher, diagnostic accuracy than other large networks while having far fewer parameters and processing faster highlights its potential usefulness in the real world scenarios, especially in disease-endemic areas where high performance computers are not available.

5.2. Comparison with Literature

Our results are consistent with and expand on prior work that has demonstrated the promise of CNNs for malaria diagnosis. Several works have proved high accuracies for different CNN architectures [2, 3, 4, 5]. But comparing them per se is not always fair, as different data sets, pre-processing and evaluation measures are used. Our work is based on a systematic comparison of many architectures eval- uated over the same OOD dataset, with all metrics featured in evaluation. The impressive results of MobileNetV2 in our work are consistent with those in other studies which have recently proposed using lightweight models for medical imaging, especially for mobile/edge deployment [12]. It is Interesting that MobileNetV2, despite its compact size compared to other larger models can outperform those in this setting and it emphasises that the feature extraction and representation learning capabilities of MobileNetV2 are well optimized for malaria picture classification specifically.

5.3. Implications of Findings

This finding has a significant impact on clinical practice and public health interventions for the control of malaria. A cost-saving and fast AI diagnostic tool by using a computational-efficient optimal CNN such as MobileNetV2 has been proved to be feasible. Tools of this nature could greatly decrease dependence on skilled microscopists, relieve the burden on health systems in endemic regions and allow for large-scale screening efforts and early detection. This could result in more timely treatment, contribute toward lowering disease transmission and ultimately to a decline in the number of patients suffering from morbidity and mortality related to malaria. In addition, the performance of MobileNetV2 implies that it could be embedded in portable diagnostic devices or mobile applications for making advanced diagnostics more accessible at/near point-of-care settings, even in remote or underserved areas.

5.4. Limitations of the Study

This study, however, has a number of limitations of consideration even with this rigorous methodology. Although the training data is regarded as a widely accepted and balanced dataset, it focuses on thin blood smear images. In future studies, these architectures could be assessed for thick blood smear images which present varied challenges with cellular debris and overlapping parasites. The investigation was also designed for two-class (infected versus uninfected) only. As future work, the approach can be extended for multi-class classification (i.e., P. falciparum and P. vivax as different classes, or Plasmodium inhuman versus other life stages) which is very important when providing appropriate guidance to follow-up treatments by specific regimen as previously obtained with sensitive detection method based on clinical data [14]. The interpretability of these models (which is briefly introduced, although incompletely), was not the main concern of this performance assessment. Additional studies using explainable AI (XAI) methods may be able to shed more light on the decision mechanisms of these CNNs, thereby increasing trust and clinical applicability.

5.5. Future Research

Based on the findings obtained, there are a number of implications for future research. The potential of ensemble methods, integrating the capabilities of multiple CNN architectures, could have been examined as well and may achieve further increased diagnostic accuracies. Second, in this context, it is also important to study which will be the impact of different augmentation strategies on how well models generalize (and potentially overfit) since there might be a data scarcity problem in these tasks. Finally, the real-time edge-device deployable solutions developed and compared based on the optimal CNN architecture found in this work would be an important advancement towards clinical implementation. Third, prospective clinical validation studies in diverse field settings would be essential for evaluating the real-life utility and impact of these AI-based diagnostic tools on malaria control initiatives.

6. Conclusion

This study inroduced a rigorous performance comparison of diverse Convolutional Neural Network (CNN) models applied to the task of automatic malaria detection from microscopic blood cell images. The main scope was to find an optimal CNN model with both high diagnostic accuracy and practical superiority including great amenity for computation in resource-limited situations. Through our comprehensive comparative study on the benchmark dataset, we found that MobileNetV2 consistently achieves state-of-the-art performance over other widely used architectures in a broad range of important diagnostic metrics such as accuracy, sensitivity, specificity, precision, F1-score and AUC. Additionally, the computational efficiency of MobileNetV2 was far better with shorter training and test times, as well much lower amount of trainable parameters making it more realistic to use in practical applications.

The implications are profound, clearly showing the huge potential for deep learning to transform malaria diagnosis with a scalable, high-throughput and unbiased solution at the point of care compared to traditional manual microscopy. The identified optimal MobileNetV2 architecture will serve as a strong basis for development of next-generation AI-based diagnostic tools, with the potential of deploying inside portable and/or mobile device which can extend its reach to remote and under-privileged populations. This development has the potential to allow no later diagnosis, earlier treatment of cases diagnosed and finally indirectly contribute to decreasing morbidity and mortality from malaria worldwide.

This study has generated a number of valuable insights, but also raises leads for future work such as multi-species classification (multi-pathogen), explainable AI methods so that the clinical relevance and interpretation can be accessed and also further validation in prospective clinical settings across multiple fields. However, this study contributes to the emerging literature on AI and global health by suggesting it may pave the way to more effective, cheap and people-centred approaches for detecting malaria that can have a dramatic influence on public health outcomes.

7. References

[1] World Health Organization. (2023). *World Malaria Report 2023*. Retrieved from https:// www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023

[2] Rajaraman, S., Antani, S. K., & Thoma, G. R. (2019). Deep learning for malaria image analysis. *Medical Image Analysis*, *57*, 233-248. Retrieved from https:// pubmed.ncbi.nlm.nih.gov/31326620/

[3] Siłka, W., Wieczorek, M., Siłka, J., & Woźniak, M. (2023). Malaria Detection Using Advanced Deep Learning Architecture. *Sensors*, *23*(3), 1501. Retrieved from https:// www.mdpi.com/1424-8220/23/3/1501

[4] Rinky, H. K., Bhuiyan, R. R., & Rahman, H. T. (2020). Performance comparison of CNN architectures for detecting Malaria diseases. In *2020 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)* (pp. 1-4). IEEE. Retrieved from https://dspace.bracu.ac.bd/xmlui/handle/10361/14725

[5] Narayanan, B. N., Ali, R., & Hardie, R. C. (2019). Performance analysis of machine learning and deep learning architectures for malaria detection on cell images. In *Applications of Machine Learning 2019* (Vol. 11139, pp. 111390W). SPIE. Retrieved from https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11139/111390W/ Performance-analysis-of-machine-learning-and-deep-learning-architectures-for/ 10.1117/12.2524681.short

[6] Rajaraman, S., Antani, S. K., & Thoma, G. R. (2018). *Malaria Cell Images Dataset*. National Library of Medicine. Retrieved from https://ceb.nlm.nih.gov/repositories/ malaria-datasets/

[7] Setiawan, A. W. (2025). Perbandingan Kinerja Arsitektur Convolutional Neural Network Pada Deteksi Malaria Menggunakan Citra Sel Darah. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, *11*(3), 643-652. Retrieved from https://jtiik.ub.ac.id/ index.php/jtiik/article/view/8085

[8] Ahamed, M. F. A., Nahiduzzaman, M., Mahmud, G., Shafi, F. B., Ayari, M. A., Khandakar, A., ... & Islam, S. M. R. (2025). Improving Malaria diagnosis through interpretable customized CNNs architectures. *Scientific Reports*, *15*(1), 1-16. Retrieved from https:// www.nature.com/articles/s41598-025-90851-1

[9] Boit, S., & Patil, R. (2024). An Efficient Deep Learning Approach for Malaria Parasite Detection in Microscopic Images. *Diagnostics*, *14*(23), 2738. Retrieved from https:// www.mdpi.com/2075-4418/14/23/2738

[10] Ikerionwu, C., Ugwuishiwu, C., Okpala, I., James, A., & Ogbodo, E. (2022). Application of machine and deep learning algorithms in optical microscopic detection of Plasmodium: A malaria diagnostic tool for the future. *Photodiagnosis and Photodynamic Therapy*, *39*, 103009. Retrieved from https://www.sciencedirect.com/ science/article/pii/S1572100022004847

[11] Kaboré, K. K., Guel, D., & Somda, F. H. (2025). Advancements in Deep Learning for Malaria Detection: A Comprehensive Overview. *International Journal of Advanced Computer Science and Applications*, *16*(1). Retrieved from https:// search.ebscohost.com/login.aspx? direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=2158107X&AN=186513893&h=ph59

[12] Setyawan, D., Wardoyo, R., & Wibowo, M. E. (2021). Malaria classification using convolutional neural network: a review. In *2021 International Conference on Informatics and Computing (ICIC)* (pp. 1-6). IEEE. Retrieved from https://ieeexplore.ieee.org/ abstract/document/9632998/

[13] Rajaraman, S., Antani, S. K., & Thoma, G. R. (2018). *Malaria Cell Images Dataset*. National Library of Medicine. Retrieved from https://ceb.nlm.nih.gov/repositories/ malaria-datasets/

[14] Asif, H. M., Khan, S. H., Alahmadi, T. J., Alsahfi, T., & Mahmoud, A. (2024). Malaria parasitic detection using a new Deep Boosted and Ensemble Learning framework. Complex & Intelligent Systems. https://doi.org/10.1007/s40747-024-01406-2

[15] Computer-aided Diagnosis of Malaria through Transfer Learning using the ResNet50 Backbone. (2023). https://doi.org/10.48550/arxiv.2304.02925

[16] Ali, S., Abdulqadir, P. S., Abdullah, S., & Yunusa, H. (2024). M2ANET: Mobile Malaria Attention Network for efficient classification of plasmodium parasites in blood cells. https://doi.org/10.48550/arxiv.2405.14242

[17] Jabbar, M., & Radhi, A. M. (2022). Diagnosis of Malaria Infected Blood Cell Digital Images using Deep Convolutional Neural Networks. Iraqi Journal of Science, 380–396. https://doi.org/10.24996/ijs.2022.63.1.35

[18] Enhancing malaria detection and classification using convolutional... (2025). https://link.springer.com/article/10.1007/s42452-025-06704-z