Malaria Detection with Lightweight DNNs:
A Study for Resource-Limited Settings

Capuzzo Daniele [1],
Trevisan Gabriele [1]**, Parussolo Elia** [1]**,**

[1]Università degli studi di Padova, Dipartimento di Ingegneria dell'Informazione

Abstract

This paper presents two lightweight DNN architectures for automated malaria detection in thin blood smear images, tailored for deployment in resource-constrained environments. Using the NIH malaria dataset, we developed and compared two approaches: a transfer learning and fine-tuning from MobileNetV2, achieving 0.967 accuracy with more than 7.4 million parameters, and an ultralight custom architecture (LightCNN) attaining a test accuracy of 0.961 with only $55,882$ parameters, and offering a viable solution for rapid diagnosis in under-resourced regions.

Code —
https://github.com/danieleCapuzzo/malaria-light-CNN
Datasets —
https://lhncbc.nlm.nih.gov/LHC-research/LHC-projects/image-processing/malaria-datasheet.html

Introduction

Malaria (Buck Emily 2023) is a life-threatening disease, transmitted via the bite of *Plasmodium* parasite infected *Anopheles* mosquitoes. It is most prevalent in subtropical and tropical regions, with sub-Saharan Africa accounting for more than 94% of the estimated cases and 95% of deaths.

According to the WHO annual malaria report (Venkatesan 2025), Malaria cases have been rising in the last few years, with the latest estimate at 263 million. The main factors contributing to such rise are the partial interruption of healthcare services during the COVID-19 pandemic, ongoing wars, and the expansion of infected mosquito habitats due to global warming.

Human diagnosis relies on the microscopic examination of blood smears by trained technicians: a process that requires minimal laboratory infrastructure, making it feasible in rural clinics, but time-consuming and labor-intensive. Moreover, human diagnosis sensitivity and specificity vary depending on technician experience and training level, ranging from 0.70 to 0.95 and 0.80 to 0.95, respectively. The previous facts make human diagnosis often unreliable, especially where microscopists cannot be properly trained.

These limitations highlight the need for automated diagnostic systems that can operate in low-resource healthcare environments. Data-driven approaches, such as Convolutional Neural Networks, have shown clear potential in computer vision tasks including automated image-based diagnosis. Indeed, CNNs are the most used architecture for state-of-the-art computer-aided diagnostic systems. However, many existing models remain infeasible to deploy on low-end devices due to the computational and dataset requirements.

In this paper we make the following contributions:

- we propose LightCNN: an ultralight custom model for automated malaria detection, which despite requiring only $55,882$ parameters, achieved an accuracy of 0.961;

- we propose an effective data augmentation pipeline designed for thin blood smear images;

- we develop and evaluate a baseline model using a transfer learning approach to provide context for our custom architecture's performance;

- we provide an extensive comparison between LightCNN, the baseline model and existing approaches in the literature, to demonstrate its effectiveness.

Related Work

The application of CNNs to malaria diagnosis has primarily focused on binary classification of red blood cells as infected or uninfected, and to a lesser extent to species-level Plasmodium parasite classification, which proved to be useful for medical treatment choice (Boggild et al. 2014).

Malaria Diagnosis (binary classification)

Numerous models have been devised to address this task:

- (Hcini, Jdey, and Ltifi 2022) introduced a custom CNN that exploits L1 regularization to reduce overfitting, achieving an accuracy of 0.997 at the cost of $3,357,090$ parameters. According to the authors, their methodology exceeded the performances of existing models.

- (Masud et al. 2020) proposed a lightweight CNN tailored to mobile applications, totaling $409,000$ trainable parameters. This model achieved a test accuracy of 0.973, obtained through dropout, batch normalization, and cyclical

learning rate. Nevertheless, Hcini found a potential risk of overfitting indicated by the distance in the training-validation curves.

- (Magotra and Rohil 2022) proposed a CNN that achieved a test accuracy of 0.96, totaling 381,729 parameters. The model outperformed both pre-trained VGG-16 and InceptionV3 (21 and 24.9 million parameters).

Other studies have evaluated and compared pre-trained networks as feature extractors. Among these, although not lightweight due to the 23 million parameters, ResNet-50 has proven successful, achieving an accuracy of 0.957.

Species-level classification

(Yang et al. 2022) developed a parasite classification pipeline where segmentation techniques are used as a pre-processing for the input of a CNN (Light-Net) which classifies the parasite among four species. Such approach achieved accuracies of 0.98 for P. Falciparum, 0.99 for P. Malariae, 0.83 for P. Ovale, and 0.84 for P. Vivax. In comparison, a VGG16-based model obtained lower accuracies of 0.96 (P. Falciparum), 0.99 (P. Malariae), 0.72 (P. Ovale), and 0.77 (P. Vivax) on the same dataset.

Gap Analysis

The research gap lies in improving the balance between diagnostic accuracy and neural network size. Although several studies have achieved impressive results in malaria detection using CNNs, most employ huge models with millions of parameters, which are impossible to deploy on the low-end devices available in rural clinics.

Methodology

For this study, we have developed two models: a custom-built lightweight CNN, referred to as LightCNN, and a model based on MobileNetV2 obtained using transfer learning and fine-tuning. MobileNetV2 demonstrated the highest performance among several evaluated models and was selected as the baseline for comparison. LightCNN was developed to achieve a comparable accuracy to the baseline, with significantly fewer parameters.

Dataset

This study utilized the publicly available NIH malaria dataset, which comprises 27,588 segmented red-blood cell images collected in Bangladesh by photographing Giemsa-stained thin blood smear slides of 200 patients, and manually labeled by an expert at the Mahidol-Oxford Tropical Medicine Research Unit, Thailand (Masud et al. 2020).
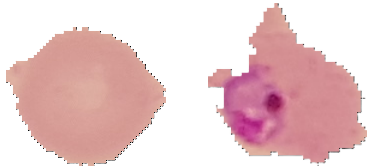

Figure 1: parasitized cells contain *Plasmodium* (right), while normal cells may contain other impurities (left)

The dataset is balanced as it contains an equal distribution of infected and uninfected cells, mitigating the risk of class imbalance leading towards predictions favoring one class over the other.

The collected RGB images vary in resolution, ranging from 46x46 to 385x395 pixels and were resized to 128×128. Pixel values were then normalized to $[0, 1]$. The dataset was partitioned 80/20 into training and test sets.

set	parasitized	uninfected	total	% dataset
training	11070	10977	22047	80%
testing	2709	2802	5511	20%

Table 1: Dataset description

The training set was further split into training and validation subsets using an additional 80/20 split. Both subsets were used for hyperparameter tuning with KerasTuner (O'Malley et al. 2019), applied to both models.

Data Augmentation

To improve model robustness and as a primary step to reduce overfitting, a data augmentation pipeline was developed and applied to the training dataset to expand it, artificially introducing images obtained from training samples by implementing a series of transformations:

- geometric, such as random rotations and flips, were utilized to introduce variations in image orientation and arrangement and a first layer of rotational invariance;
- illumination (random brightness and contrast changes) were applied to simulate lightning condition variations;
- Gaussian noise was introduced to emulate real-world camera sensor noise.

Baseline Model

The baseline model was derived by removing the original top classification layers from the pre-trained MobileNetV2. The base was retained and frozen to serve as a feature extractor. A custom multi-layer perceptron (MLP) replaced the classification head, resulting in the following architecture:

Layer (type)	Output shape	Param #
MobileNetV2	(None, 4, 4, 1280)	2,257,984
GlobalAvgPooling2D	(None, 1280)	0
BatchNormalization	(None, 1280)	5,120
Dropout	(None, 1280)	0
Dense	(None, 192)	245,952
Dropout 1	(None, 192)	0
Dense 1	(None, 1)	193

Table 2: Baseline architecture ($2,472,577$ trainable params)

Fine-tuning was subsequently performed by training the whole network for 3 epochs, using a learning rate of 10^{-4}.

LightCNN

LightCNN is an ultra-light model designed for maximal efficiency, while retaining state-of-the-art accuracy. The high degree of computational efficiency is achieved by leveraging SeparableConv2D: a layer that performs a depthwise convolution, acting separately on each channel, followed by a pointwise convolution to recombine the feature maps.

Layer (type)	Output shape	Param #
Block 1	(None, 64, 64, 32)	283
Block 2	(None, 32, 32, 32)	1,472
Block 3	(None, 16, 16, 56)	2,360
Block 4	(None, 8, 8, 88)	5,872
GlobalAvgPooling2D	(None, 88)	0
Dense+BN+Dropout	(None, 96)	8,928
Dense (output)	(None, 1)	97

Table 3: LightCNN architecture ($18,404$ trainable params)

The architecture detailed in the previous table consists of blocks, each comprising a SeparableConv2D layer, followed by BatchNormalization, MaxPooling(2,2), and Dropout to mitigate overfitting and improve generalization. Furthermore GlobalAveragePooling2D was used in place of Flatten to further reduce the model complexity.

Keras Tuner identified the following optimal hyperparameters: *Leaky ReLU*, *Adam* optimizer (*learning rate* $\simeq 0.00102$), convolutional filters, dropout rates of $(0.175, 0.0, 0.075, 0.05, 0.4)$, and the number of dense units.

Results and Discussion

In this section, we present and discuss the results obtained by our proposed models on the test dataset, which comprises 5511 images, to asses their diagnostic effectiveness and suitability for deployment in the specified settings. To this end, we employed several standard evaluation metrics.

Metric	Class	LightCNN	Baseline
Accuracy	*Overall*	0.96	0.97
MCC	*Overall*	0.92	0.93
Precision	Uninf.	**0.98**	0.97
	Parasit.	0.94	0.96
	Average	0.96	0.97
Recall	Uninf.	0.94	0.96
	Parasit.	**0.98**	0.98
	Average	0.96	0.96
F1-score	*Average*	0.96	0.97
# train. params	*Overall*	18K	2.5M

Table 4: LightCNN vs. Baseline

As can be observed from the data reported in the table above, LightCNN achieves a slightly lower overall accuracy of 0.96 and Matthew's correlation coefficient of 0.92 compared to the baseline model's 0.97 and 0.93, respectively. Although the performances may seem worse, these don't paint the whole picture, as neither accuracy nor MCC can distinguish between the relative cost of specific error types.

Although LightCNN shows 0.01 lower precision, recall and F1-score, a closer analysis reveals that it achieved higher class-based, uninfected precision (of 0.98) and higher recall for the parasitized class (of 0.98). The confusion matrices confirm that the model is less prone to false negatives, suggesting it to be more effective at detecting infected cells: this represents a critical advantage in the clinical context.

Figure 2: LightCNN predicted only 47 false negatives, as opposed to 69 of the Baseline.

The ROC curves nearly hugging the top-left corner and the close-to-one AUC-ROC scores prove that both models possess an excellent ability to discriminate between classes
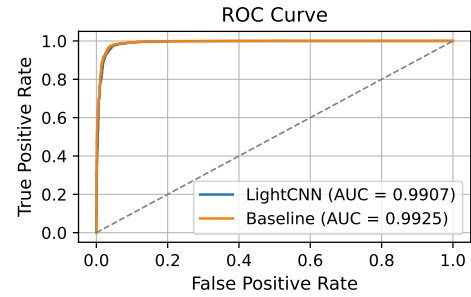

Figure 3: ROC-curves of the proposed models

Despite the baseline having a modest edge, likely due to the superior number of parameters which help in dealing with complex patterns, this gain is negligible considering its size.
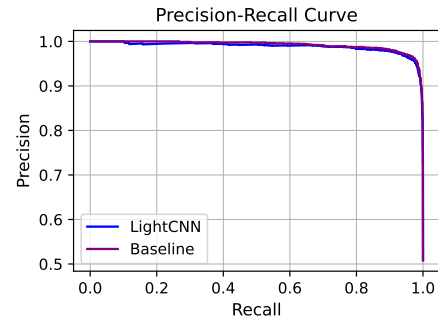

Figure 4: Precision-recall-curves of the proposed models

The precision-recall curves confirm the comparable performances of the two models, with baseline possessing a modestly higher discriminative power, at the cost of producing more false negatives than false positives. Overall the error profile of LightCNN is advantageous in medical diagnosis.

Comparison with Related Papers

To further assess the efficiency of LightCNN we conducted a comparison with the best-performing models discussed in the related work section.

Model	Acc.	AUC	Prec	Rec	#params
(Hcini)	0.99	-	-	-	3.3M*
(Masud)	0.97	0.97	0.97	0.97	409K
(Rajaraman)					381K*
custom	0.94	0.98	0.95	0.93	438K*
ResNet-50	0.96	0.99	0.97	0.95	23M**
VGG-16	0.94	0.98	0.95	0.94	138M**
(**Proposed**)					
LightCNN	0.96	0.99	0.96	0.96	18K
Baseline	0.97	0.99	0.97	0.97	2.5M

Table 5: *calculated, **unavailable #trainable parameters.

LightCNN proves to be the lightest model by a significant margin whilst closely performing at a level comparable to or $1/2\%$ worse than leading models across all metrics, even surpassing the performance of all models proposed by (Rajaraman et al. 2018) in every metric, except precision.

Even the proposed baseline model, despite its larger size, proves to be lighter than most models currently in research, and it consistently outperforms them in all criteria, with the only exception being (Hcini, Jdey, and Ltifi 2022) accuracy.

Final considerations

The performed analysis conclusively demonstrates LightCNN's exceptional classification performance in spite of its remarkably small size, making it the most suitable model for deployment in resource-constrained settings, where low inference time and FLOPs are critical measures to take into account.

Model	FLOPs	Avg. Inference Time (ms)
lightCNN	24,785,153	13.10
baseline	200,645,121	97.68

Table 6: LightCNN's inference time is more than 7 times faster and requires significantly fewer FLOPs.

Social Impact and Ethical Considerations

The employment of fully automated malaria detection tools, such as LightCNN, can have a positive impact on communities in underserved regions of the world, where access to properly trained medical professionals is limited. LightCNN's efficiency enables the deployment on low-end hardware, facilitating accurate diagnosis where it is needed the most. From an economic perspective, early diagnosis and treatment reduces the progression to severe illness, decreasing the necessity of specialized equipment and expensive medical care, thereby lowering strain on hospitals, and ultimately securing better and timely treatment for grave patients.

LightCNN also offers an affordable and rapid method for real-time malaria screening, providing crucial data on disease diffusion and facilitating the implementation of targeted preventive health policies aimed at interrupting the transmission cycle, effectively reducing the spread of the disease.

Ethical Implications

Although AI laws vary from country to country (Basu and Dave 2025), the deployment of deep-learning-based diagnostic systems must be conducted in compliance with proper ethical medical standards, prioritizing strict adherence to laws on confidentiality and privacy to protect sensitive patient data and according to the ethical principles of informed consent (Shah et al. 2025), patients should be informed about the potential use of AI in the diagnostic process.

Furthermore, transparent documentation should be provided so that medical professionals can understand these system's capabilities and limitations, maintain clinical judgment and preserve the human element in decision-making.

Conclusions and Future Work

In this study, we proposed and evaluated two custom deep-learning models for automated malaria detection, focusing on deployability in resource-constrained environments to achieve a the highest possible potential social impact. To this end, the developed ultra-light architecture LightCNN demonstrated both performance comparable to state-of-the-art models, and exceptional efficiency thanks to our data augmentation pipeline tailored to medical settings and our use of KerasTuner.

LightCNN's combination of 96.1% accuracy, with 130 times fewer parameters, more than 7x improvement in inference and FLOPs, compared to the baseline, resulting in lower memory and energy consumption, highlight its usability in the specified settings, even in battery-powered devices where the electrical grid is not reliable.

Future Work

This study lays a possible groundwork for further research in the field of lightweight medical neural networks, needed to assess flaws and necessary improvements, specifically:

- ensuring clinical reliability by evaluating the model on multiple datasets representing both different geographical regions and patient demographics;

- the integration of explainable AI techniques to clarify the decision process and the addition of human-aided decision tools for uncertain cases;

- the implementation of online learning strategies to better adapt the model to local data and/or each clinic's specific laboratory settings. Alternatively, evaluating the possibility of collecting local data with a periodic centralized training;

- conducting comprehensive usability and ethical studies to better assess the deployability in real-world scenarios and the actual social impact.

References

Basu, N.; and Dave, R. 2025. Comparative Analysis of Laws in AI. *Journal of Lifestyle and SDGs Review*, 5: e05575.

Boggild, A.; Brophy, J.; Charlebois, P.; Crockett, M.; Geduld, J.; Ghesquiere, W.; McDonald, P.; Plourde, P.; Teitelbaum, P.; Tepper, M.; Schofield, S.; and McCarthy, A. 2014. Summary of recommendations for the diagnosis and treatment of malaria by the Committee to Advise on Tropical Medicine and Travel (CATMAT). *Can. Commun. Dis. Rep.*, 40(7): 133–143.

Buck Emily, F. N. A. 2023. Malaria. Updated July 31, 2023.

Hcini, G.; Jdey, I.; and Ltifi, H. 2022. Improving Malaria Detection Using L1 Regularization Neural Network. *JUCS - Journal of Universal Computer Science*, 28(10): 1087–1107.

Magotra, V.; and Rohil, M. K. 2022. Malaria Diagnosis Using a Lightweight Deep Convolutional Neural Network. *International Journal of Telemedicine and Applications*, 2022: 1–8.

Masud, M.; Alhumyani, H.; Alshamrani, S. S.; Cheikhrouhou, O.; Ibrahim, S.; Muhammad, G.; Hossain, M. S.; and Shorfuzzaman, M. 2020. Leveraging Deep Learning Techniques for Malaria Parasite Detection Using Mobile Application. *Wireless Communications and Mobile Computing*, 2020(1): 8895429.

O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L.; et al. 2019. KerasTuner. https://github.com/keras-team/keras-tuner.

Rajaraman, S.; Antani, S.; Poostchi, M.; Silamut, K.; Hossain, M.; Maude, R.; Jaeger, S.; and Thoma, G. 2018. Pretrained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, 6.

Shah, P.; Thornton, I.; Kopitnik, N.; and et al. 2025. Informed Consent. *StatPearls [Internet]*. [Updated 2024 Nov 24].

Venkatesan, P. 2025. WHO world malaria report 2024. *Lancet Microbe*, 6(4): 101073.

Yang, Z.; Benhabiles, H.; Hammoudi, K.; Windal, F.; He, R.; and Collard, D. 2022. A generalized deep learning-based framework for assistance to the human malaria diagnosis from microscopic images. *Neural Comput. Appl.*, 34(17): 14223–14238.