Inteligencia Artificial 28(76) (2025), 186-199
doi: 10.4114/intartif.vol28iss76pp186-199

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

M^2ANET: Mobile Malaria Attention Network for efficient classification of plasmodium parasites in blood cells

Salam Ahmed [1], Peshraw Salam Abdalqadir [2], Shan Ali Abdullah [2], *Yunusa Haruna [3]

[1] Department of Computer Science, University of Garmian, Sulaymaniyah-Kalar, Iraq
[2] Department of Computer Science, Darbandikhan Technical Institute, Sulaimani Polytechnic University, Sulaimani-Iraq.
[3] School of Automation Science and Electrical Engineering, Beihang University, Beijing, China

[1] salam.ahmed@garmian.edu.krd
[2] peshraw.s.abdalqadir@spu.edu.iq, shankaloshyali1999@gmail.com
[3] yunusa2k2@buaa.edu.cn

Abstract Malaria is a life-threatening infectious disease caused by Plasmodium parasites, which poses a significant public health challenge worldwide, particularly in tropical and subtropical regions. Timely and accurate detection of malaria parasites in blood cells is crucial for effective treatment and control of the disease. In recent years, deep learning techniques have demonstrated remarkable success in medical image analysis tasks, offering promising avenues for improving diagnostic accuracy. However, limited studies focus on hybrid mobile models due to the complexity of combining two distinct architectures and the significant memory demand of self-attention mechanisms, especially for edge devices. In this study, we introduce M^2ANET (Mobile Malaria Attention Network), a hybrid model integrating MBConv3 (MobileNetV3 blocks) for efficient local feature extraction and a modified global-MHSA (multi-head self-attention) mechanism for capturing global context in blood cell images. Experimental results on the Malaria Cell Images Dataset show that M^2ANET achieves a top-1 accuracy of 95.45% and a Cohen Kappa score of 0.91, outperforming some state-of-the-art lightweight and mobile networks. These results highlight its effectiveness and efficiency for malaria diagnosis. The development of M^2ANET demonstrates the potential of hybrid mobile models for improving malaria diagnosis in resource-constrained settings.

Keywords: Attention mechanism, Malaria detection, Medical image analysis, Mobile hybrid model.

1 Introduction

The diagnosis and detection of malaria, a life-threatening infectious disease caused by Plasmodium parasites transmitted through the bites of infected female mosquitoes, remain critical challenges in global healthcare [1]. With millions of cases reported annually, particularly in tropical and subtropical regions, the timely and accurate identification of malaria parasites in blood cells is crucial for effective treatment and control [2, 3].

However, the current method of malaria diagnosis relies on manual microscopic examination of the appearance, number, and shape of red blood cells. This approach necessitates the involvement of skilled medical experts, rendering the process both time-consuming and costly [4, 5, 6, 7]. Moreover, the subjective nature of visual examination means that diagnostic results may occasionally be inaccurate due to human errors, highlighting the need for more reliable and efficient automated visual diagnostic method [8].

In recent years, computer vision techniques have demonstrated remarkable success in various medical image analysis tasks, including disease detection and diagnosis [9]. However, these methods are primarily deployed as computer-aided diagnostic (CAD) systems to provide rapid assistance and enhance the accuracy of disease diagnosis [10]. Previous research has focused on classifying and detecting malaria in cells using conventional methods such

ISSN: 1137-3601 (print), 1988-3064 (on-line)

as Convolutional Neural Network (CNN) architectures like LeNet [11], which has a shallow depth and limited feature extraction capabilities, making it less effective for complex tasks, VGG [12], which has a large number of parameters and computational complexity, and issues like vanishing gradients, making it inefficient for resource-constrained environments, AlexNet [13], which still faces challenges in terms of computational demand and generalization, Inception [14], ResNet [15], and EfficientNet [16], which improved on computational complexity and memory efficiency, allowing for deeper networks but still struggle to capture the global con-text of an image due to their strong inductive bias, as well as other non-CNN methods like Support Vector Machine (SVM) [17] and XG-Boost [18, 19], which are less suitable for image classification tasks as they do not directly learn spatial features from raw pixel data like CNNs do. While these approaches have shown good performance, the use of custom methods designed for specific disease detection tasks has the potential for greater precision, efficiency, and reliability. Moreover, limited research has explored the utilization of self-attention mechanisms to enhance malaria disease detection.

The self-attention mechanism was initially developed for natural language processing but later adapted for computer vision models to enhance capturing long-range dependencies and relationships within image input data [20]. This mechanism allows the model to dynamically focus on relevant regions of the image while considering the interactions between different parts of the input. It helps the model to understand the global context of an image, which is particularly beneficial when identifying complex patterns that require holistic analysis rather than just local features [21]. In medical image analysis, self-attention is especially important because it allows for more accurate identification of subtle and dispersed features within medical images, such as the irregular shapes and sizes of Plasmodium parasites in blood smears. By considering the entire image context, self-attention can enhance the detection of anomalies that might be missed by models relying solely on local information.

However, the application of self-attention in vision models, particularly in resource-constrained environments like mobile and edge devices, faces significant challenges. Self-attention mechanisms are computationally intensive and memory-demanding, which limits their practicality in these settings [22]. Hybrid models, which combine CNNs with self-attention mechanisms, offer a promising solution to these limitations. By leveraging the efficient local feature extraction capabilities of CNNs and integrating them with the global context capturing power of self-attention, hybrid models can achieve a balance between computational efficiency and performance [21]. This approach ensures that the model remains lightweight and suitable for deployment on mobile and edge devices while maintaining high accuracy and robustness in medical image analysis. Figure 1 shows the performance of M^2ANET in classifying Plasmodium parasites in blood cell images compared to some recent methods in terms of accuracy and trainable parameters.

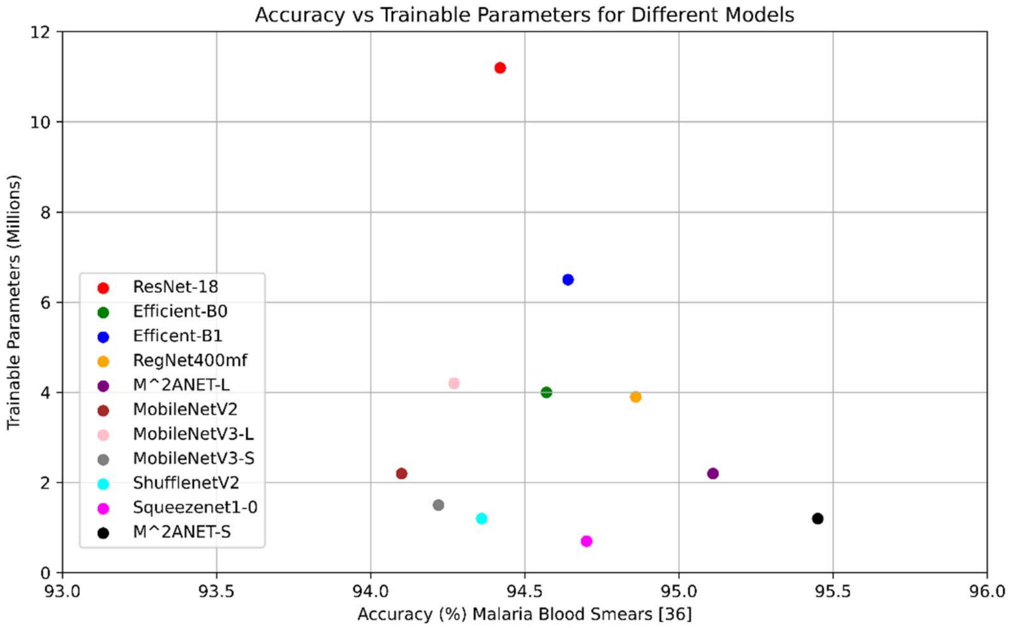


Figure 1 M^2ANET comparison with recent methods

In this study, we design a novel model for detecting malaria in blood cells that classifies parasitized and unparasitized cells. The model is a hybrid mobile network using MBConv3 [23] in the first two stages for efficient local features extraction and modified 2D MHSA (multi-head self-attention) in the latter stages for improved global context captures. We fused both layers using a pair-wise addition. The network is dubbed as M^2ANET (Mobile Malaria Attention Network). M^2ANET integrates attention mechanisms to dynamically highlight informative regions within cell images while maintaining computational efficiency. This integration allows M^2ANET to surpass conventional mobile models like MobileNets [23], ShuffleNet [24], SqueezeNet [25], etc. while remaining computationally efficient, making it suitable for deployment in resource constrained mobile and edge devices such as healthcare facilities in malaria endemic regions. This study represents a significant step forward in the search for efficient, accurate and low-resource models for malaria disease detection, aiming to accurately diagnose this deadly disease.

The contributions of this paper are summarized as follows:

- We propose M^2ANET, an efficient mobile-based hybrid model for detecting malaria disease using red blood cells images.
- The model is designed for mobile and edge devices which is computationally efficient in real time.
- The model can serve as a baseline hybrid for identifying plasmodium parasites in blood cells images, where developers and researcher can continue to explore the synergy of employing two distinct models efficiently.

2 Related Work

Deep Learning Methods. Previous studies have been conducted to detect malaria in blood cells with promising results using deep learning methods including Deep Belief Network (DBN) and CNN. Bibin et al. (2017) [26] proposed a novel method utilizing DBN for detecting malaria parasites in peripheral blood smear images, achieving high F-score of 89.66% through pre-training with contrastive divergence. While, Sivaramakrishnan et al. (2017) [27] proposed a custom CNN model for malaria cell classification, achieving a high accuracy of 98.61% by visualizing features and activations within the model. Then again, Sivaramakrishnan et al. (2018) [28] evaluated pre-trained CNNs for malaria parasite detection in blood smear images, achieving sensitivity and specificity scores of 0.992 and 0.986 respectively, for feature extraction and classification. Yang et al. (2020) [29] developed a deep learning method for automated malaria parasite detection in thick blood smear images using smartphones, achieving a lower accuracy of 93.46% due to sacrificing accuracy to computational complexity. Vijayalakshmi et al. (2020) [30] introduced a deep neural network model for identifying infected falciparum malaria parasites using transfer learning with VGG-SVM, outperforming existing CNN models in accuracy and performance indicators. Loddo et al. (2022) [31] investigated deep learning architectures for malaria diagnosis, comparing conventional CNN models and evaluating their performance on different datasets, highlighting the need for further research to improve robustness, though their study achieved an accuracy of 95.2%. Madhu et al. (2021) [32] developed a Deep Siamese Capsule Network (D-SCN) model for automatic diagnosis of malaria parasites in thin blood smears, achieving high detection scores of 97.24% and classification accuracy scores of 98.89%. Siłka et al. (2023) [33] proposed an AI-based object detection system for malaria diagnosis, achieving 99.68% accuracy comparable to human microscopists, which could aid diagnosis in resource-limited regions. Abdurahman et al. (2021) [34] investigated modified YOLOV3 and YOLOV4 models for malaria parasite detection in thick blood smear images, achieving a mAP score of 96.14% and 95.46% respectively, outperforming other state-of-the-art detection methods. Zhong et al. (2023) [35] developed an efficient malaria detection system using CNN adapted for mobile devices and microscopes, achieving 97.74% accuracy and 97.75% F-score with diverse image dataset of various regions.

Machine Learning Methods. Aris et al. (2020) [36] proposed a fast k-means clustering algorithm for malaria detection in thick blood smear images, evaluating 5 color models and 15 color components. Their study concludes that segmentation through the R component of RGB achieves the highest accuracy at 99.81%. Jahan & Alam (2023) [37] in-troduce a hybrid machine learning algorithm to classify malaria-infected erythro-cytes, combining supervised algorithms such as stochastic gradient descent, lo-gistic regression, decision trees, and XGBoost. Python-based approach achieves 95.64% and 96.22% accuracy in two configurations, aiding medical practitioners in malaria diagnosis. Murmu & Kumar (2024) [38] present DLRFNet, combines deep CNN with Random Forest to detect plasmodium malaria parasite. Their method address data scarcity and imbalance, improving on existing models and shows it effectiveness in parasite detection, leverages domain-specific expertise and integrating modifications for enhanced visualization and precise boundary detection.

Discussion. The review highlights a notable lack of exploration into self-attention mechanisms and hybrid mobile model designs for malaria detection. Self-attention mechanisms enable models to capture global context within images, addressing the limitations of CNNs that primarily focus on local details. While CNNs excel in lightweight and efficient feature extraction, they struggle to effectively interpret complex spatial dependencies across entire images. Hybrid mobile designs, which integrate lightweight convolutional operations with self-attention mechanisms, present a solution by leveraging the strengths of both approaches. This combination ensures efficient processing while maintaining high accuracy, making it particularly suitable for deployment in resource-constrained environments such as mobile and edge devices.

Furthermore, as discussed in the Introduction and Related Work, existing models often exhibit high computational costs and significant memory demands, rendering them impractical for real-time diagnostics in healthcare facilities, especially in malaria-endemic regions. By effectively combining CNN for localized feature extraction and self-attention for global context understanding, our proposed M^2ANET addresses these challenges. The model ensures precise classification of parasitized and un-parasitized red blood cells while being computationally efficient and deployable on mobile and edge devices. This work provides a significant contribution by bridging the gap in current methodologies, offering a practical and scalable solution for improving malaria diagnosis in resource-constrained settings**.**

3 Method

3.1 M^2ANET Overview

M^2ANET is a novel hybrid mobile deep learning model designed to enhance the classification accuracy of blood cell images, particularly in classifying between plasmodium parasitized and non-parasitized cells. It achieves this through the combination of spatial feature extraction capabilities from MBConv3, based on the MobileNetV3 architecture, and a modified 2D global MHSA. The modification introduces a grouped point-wise convolution to the query, key and value projections in the MHSA, effectively reducing computational and memory complexity. This integration facilitates efficient processing in resource-constrained environments, such as mobile and edge devices, while precisely capturing local and global context within blood cell images, thereby enhancing the reliability and accuracy of malaria diagnosis. Figure 2 shows architectural design of the model. Figure 2 illustrates the architectural design of the model.

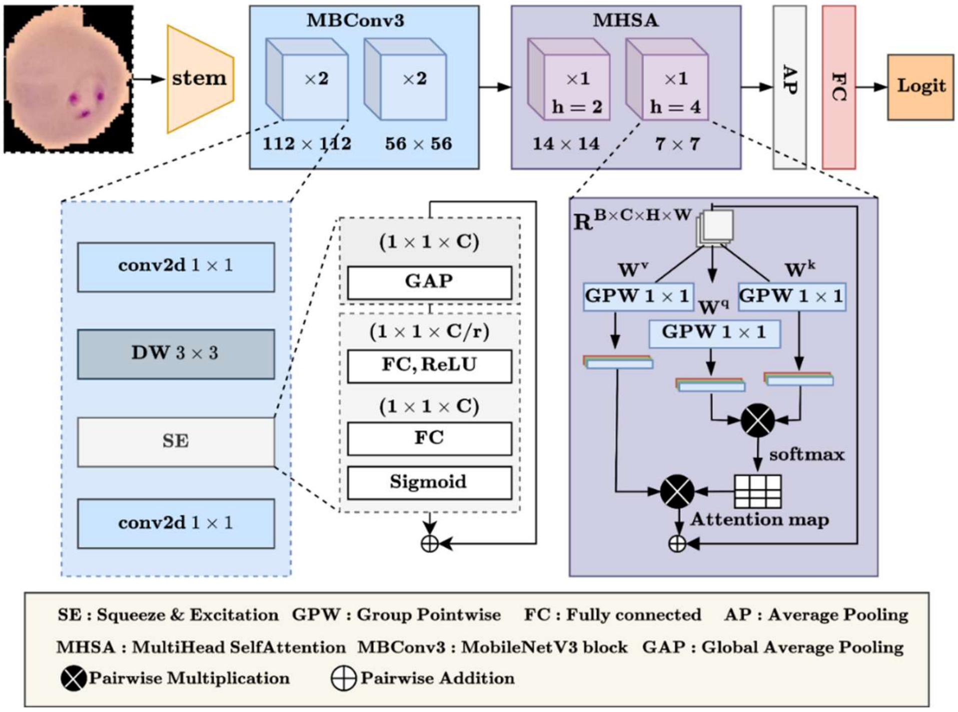


Figure 2 M^2ANET architecture

3.2 M^2ANET Architectural Design

Input. M^2ANET processes 2D RGB images of size 112 × 112 pixels, with each image having to encode the color information, enabling the model to recognise visual features and patterns. The consistent small image size ensures uniform processing, and within the network and facilitates compatibility across various low-resourced edge devices. Standardizing the input dimensions allows to effectively analyze and extract relevant information, enabling efficient robust plasmodium parasite cell disease detection.

Stem. The stem block in M^2ANET serves as a basic feature extractor, aiming to extract fine-grained information from the input image before reducing its spatial dimensions. We preferred this approach over directly reducing the spatial dimension for computational efficiency, as the latter leads to information loss. The stem block is one layer of 3×3 convolution with batch normalization and ReLU activation, all with a stride of 1. Let X denote the input Plasmodium cell image, where $X \in \mathbb{R}^{H\times W\times C}$.

Efficient Local Details. The section of this network aims to extract fine-grained local-features that is low on computational demand which will efficiently run on edge-devices and mobile devices. Therefore, we utilized the MBConv3 blocks (MobileNetV3) to serve this purpose since they are computationally efficient and built to work in mobile and edge devices. These blocks adopt a bottleneck design with blocks, a pairwise conv 1×1 for projection, then a 3×3 depth-wise separable convolution, then a squeeze & excitation (SE) layer for channel-wise calibration, and lastly a 1×1 conv for dimensionality. For our design we arranged these blocks as [2, 2] for local-details extraction. We represent the operations within the MBConv3 block from equation 1- 4.

Pairwise convolution (1×1 convolution)

$$\mathrm{Proj(x)} = \mathrm{Conv}_{1\times1}\left(\mathrm{x}, \mathrm{W_{proj}}\right) + \mathrm{b_{proj}} \quad (1)$$

Depth-wise separable convolution

$$\mathrm{DWConv(x)} = \mathrm{Conv}_{3\times3}^{\mathrm{depth-wise}}(\mathrm{x}, \mathrm{W_{DW}}) + \mathrm{b_{DW}} \quad (2)$$

Squeeze and Excitation (SE) Layer

$$\mathrm{SE(x)} = \sigma(\mathrm{avgpool}\left(\mathrm{ReLU}\left(\mathrm{Conv}_{1\times1}\left(\mathrm{x}, \mathrm{W_{squeeze}}\right) + \mathrm{b_{squeeze}}\right)\right) \odot \mathrm{x} \quad (3)$$

Dimensionality reduction

$$\mathrm{DimRed\ (x)} = \mathrm{Conv}_{1\times1}(\mathrm{x}, \mathrm{W_{dimred}}) + \mathrm{b_{dimred}} \quad (4)$$

Lightweight Global Details. This section discusses the integration of 2D *global* MHSA into M^2ANET, emphasizing the need to reduce computational complexity for mobile devices. In the conventional approach, pointwise convolutions are applied to the query, key, and value projections within MHSA which is effective but can lead to significant computational overhead, particularly when handling high-dimensional inputs. To address this, we introduce the use of grouped pointwise convolutions, which maintain performance while substantially improving computational efficiency.

Grouped pointwise convolutions involve applying convolutions independently to groups of channels within the input tensor. This approach reduces the number of parameters and the overall computational complexity. In the standard configuration, each input channel interacts with every output channel, resulting in a computational burden. Specifically, for input tensor $X \in \mathbb{R}^{n\mathrm{x}c\mathrm{x}\mathrm{h}\mathrm{x}\mathrm{w}}$ with c channels, the pointwise convolution is applied as in equation 5.

$$Y = X * W + b \quad (5)$$

where, $W \in \mathbb{R}^{c\mathrm{x}c\mathrm{x}1\times1}$ and $b \in \mathbb{R}^{\mathrm{c}}$

To enhance efficiency, we modify the convolutions to be grouped. Each group processes a subset of the input channels independently. Given the same input tensor, the grouped convolution is applied with g groups (where $g = c$), such that each group contains one channel as in equation 6.

$$Y = X * W + b \quad (6)$$

Where, $W_g \in \mathbb{R}^{(^C/_g)\times(^C/_g)\times1\times1}$ and $b_g \in \mathbb{R}^{^C/_g}$

The choice of grouped pointwise convolutions is due to their ability to decouple the interactions between different channels. By processing each channel independently, the computational complexity is reduced from $\mathrm{O}(C^2)$

to $\mathrm{O}(C)$, where C is the number of channels. This reduction in complexity translates to fewer parameters and operations, thereby improving computational efficiency while maintaining the expressiveness of the model. Thus, enhances the model's suitability for deployment on mobile devices. Despite the streamlined computations, the proposed method ensures that the performance of the attention mechanism remains robust. This balance between efficiency and performance is crucial for mobile classifiers that need to operate under resource constraints environment without compromising accuracy.

Non-isotropic architecture. Since M^2ANET is a hybrid model that integrates an attention mechanism which mostly functions in an isotropic architecture (i.e., maintaining the same feature spatial resolution throughout the whole depths with no down-sampling). However, this choice of design is computationally costly because it treats all input dimensions equally, regardless of their orientation or position. This means that isotropic architecture requires a larger number of parameters to learn the same level of complexity as non-isotropic architecture. Therefore, we adopted the pyramid-like structure of traditional CNNs. M^2ANET down-samples feature maps by applying a stride of 2 after each stage of the network, thus down-sampling the feature maps and significantly reducing the model's computational complexity, making it suitable for low-resource environments.

Interaction between Local and Global Features. $\boldsymbol{M^2}$ANET achieves a seamless fusion of local and global features within its architecture. This fusion involves combining the outputs from the MBConv3 blocks, representing local features ($\boldsymbol{L_{local}}$), with the inputs to the 2D global MHSA, representing global features ($\boldsymbol{G_{global}}$), equation 7.

$$F_{fused} = L_{local} \oplus G_{global} \tag{7}$$

Here, $\oplus$ denotes the fusion operation. This integration mechanism ensures that both fine-grained details and broader contextual information are effectively combined to achieve a comprehensive understanding of the input image

4 Experimental Results and Comparison

This section presents the experimental results and comparisons with state-of-the-art mobile networks such as MobileNetV2 [23], MobileNetV3-L [23], MobileNetV3-S [23], ShufflenetV2 [24], Squeezenet1-0 [25], as well as lightweight models such as ResNet-18 [40], Efficient-B0 [41], Efficent-B1[41] and RegNet400mf [42]. The models were evaluated using malaria-infected thin blood smear images to classify parasitized and non-parasitized cells.

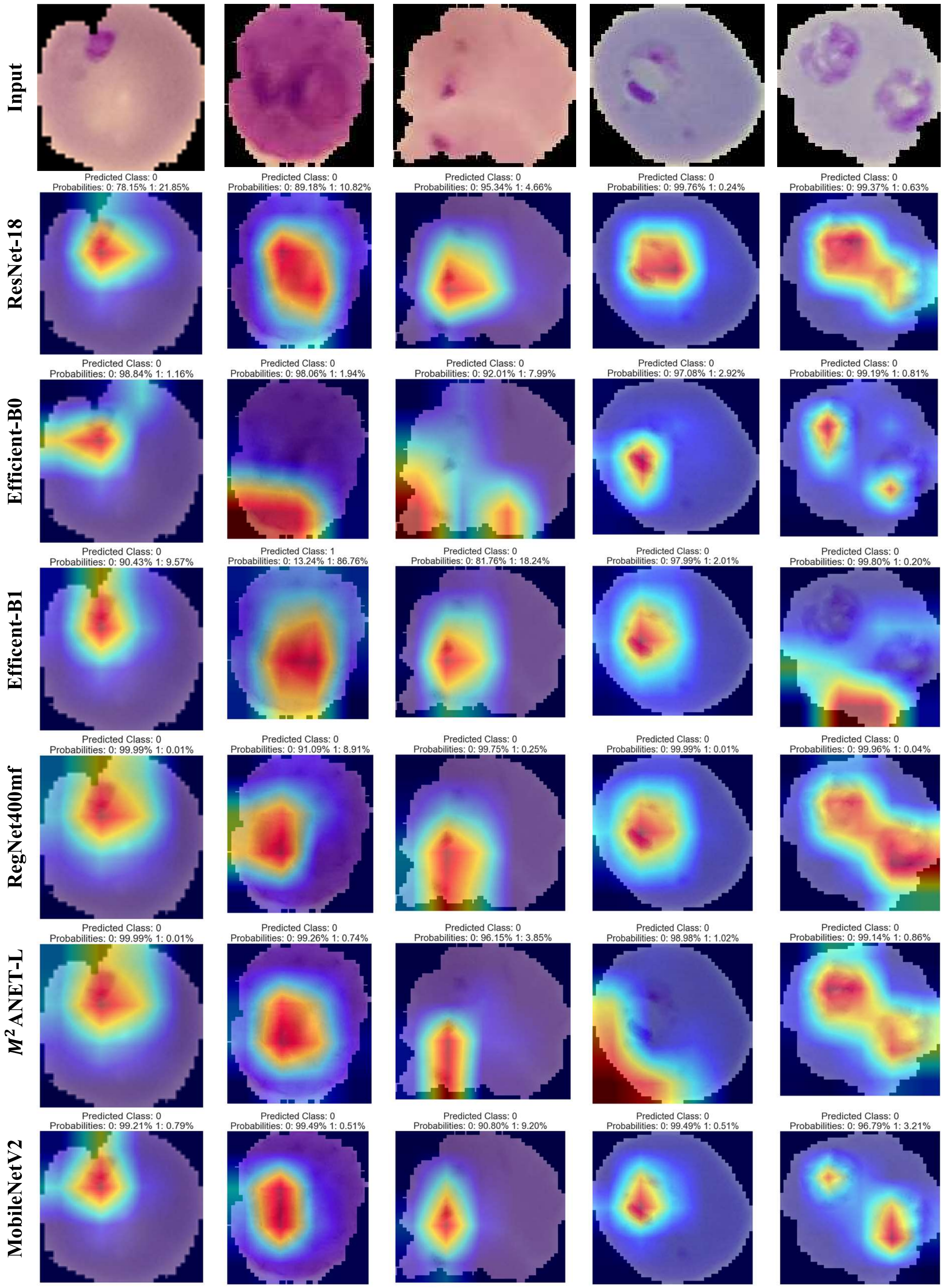
Datasets. The Malaria dataset comprises of 27,558 cell images, ranging from 150 to 150 pixels, evenly divided between 13,779 parasitized cell images and 13,779 uninfected cells images. These images are derived from thin blood smear slides of segmented cells [39].

Experimental Configuration. The experiments were performed on a Linux-based machine equipped with an Intel Core i7 8700k processor, two NVIDIA Titan XP 12GB GPUs, and 32GB of RAM. The training procedure included 90 epochs with a batch size of 64. The AdamW optimizer was used with an initial learning rate of 0.0001 and a weight decay rate of 0.05.

4.1 Visual Explanations with Grad-CAM

To assess and interpret the decision-making process of M^2ANET in detecting Plasmodium parasites from thin blood smear images, we employed Gradient-weighted Class Activation Mapping (Grad-CAM) [43]. Grad-CAM is a widely used explainable AI (XAI) technique that highlights the important regions in an image that contribute most to the model's prediction. By generating heat-maps over the input images, we visualized the spatial attention of M^2ANET, confirming its ability to localize infected regions with better precision. These visualizations provide transparent insights into the internal workings of the model, helping validate its predictions and ensuring they are not based on false patterns or irrelevant features.

The interpretability offered by Grad-CAM is essential in medical imaging applications, where trust and clarity in model decisions are critical. Figure 3 presents a comparison of Grad-CAM results across M^2ANET and several state-of-the-art baselines, illustrating that M^2ANET not only achieves better performance metrics but also demonstrates consistent attention to plasmodium parasite relevant regions in the blood cells.


Input
ResNet-18
Predicted Class: 0
Probabilities: 0: 78.15% 1: 21.85%
Predicted Class: 0
Probabilities: 0: 89.18% 1: 10.82%
Predicted Class: 0
Probabilities: 0: 95.34% 1: 4.66%
Predicted Class: 0
Probabilities: 0: 99.76% 1: 0.24%
Predicted Class: 0
Probabilities: 0: 99.37% 1: 0.63%
Efficient-B0
Predicted Class: 0
Probabilities: 0: 98.84% 1: 1.16%
Predicted Class: 0
Probabilities: 0: 98.06% 1: 1.94%
Predicted Class: 0
Probabilities: 0: 92.01% 1: 7.99%
Predicted Class: 0
Probabilities: 0: 97.08% 1: 2.92%
Predicted Class: 0
Probabilities: 0: 99.19% 1: 0.81%
Efficent-B1
Predicted Class: 0
Probabilities: 0: 90.43% 1: 9.57%
Predicted Class: 1
Probabilities: 0: 13.24% 1: 86.76%
Predicted Class: 0
Probabilities: 0: 81.76% 1: 18.24%
Predicted Class: 0
Probabilities: 0: 97.99% 1: 2.01%
Predicted Class: 0
Probabilities: 0: 99.80% 1: 0.20%
RegNet400mf
Predicted Class: 0
Probabilities: 0: 99.99% 1: 0.01%
Predicted Class: 0
Probabilities: 0: 91.09% 1: 8.91%
Predicted Class: 0
Probabilities: 0: 99.75% 1: 0.25%
Predicted Class: 0
Probabilities: 0: 99.99% 1: 0.01%
Predicted Class: 0
Probabilities: 0: 99.96% 1: 0.04%
M^2ANET-L
Predicted Class: 0
Probabilities: 0: 99.99% 1: 0.01%
Predicted Class: 0
Probabilities: 0: 99.26% 1: 0.74%
Predicted Class: 0
Probabilities: 0: 96.15% 1: 3.85%
Predicted Class: 0
Probabilities: 0: 98.98% 1: 1.02%
Predicted Class: 0
Probabilities: 0: 99.14% 1: 0.86%
MobileNetV2
Predicted Class: 0
Probabilities: 0: 99.21% 1: 0.79%
Predicted Class: 0
Probabilities: 0: 99.49% 1: 0.51%
Predicted Class: 0
Probabilities: 0: 90.80% 1: 9.20%
Predicted Class: 0
Probabilities: 0: 99.49% 1: 0.51%
Predicted Class: 0
Probabilities: 0: 96.79% 1: 3.21%

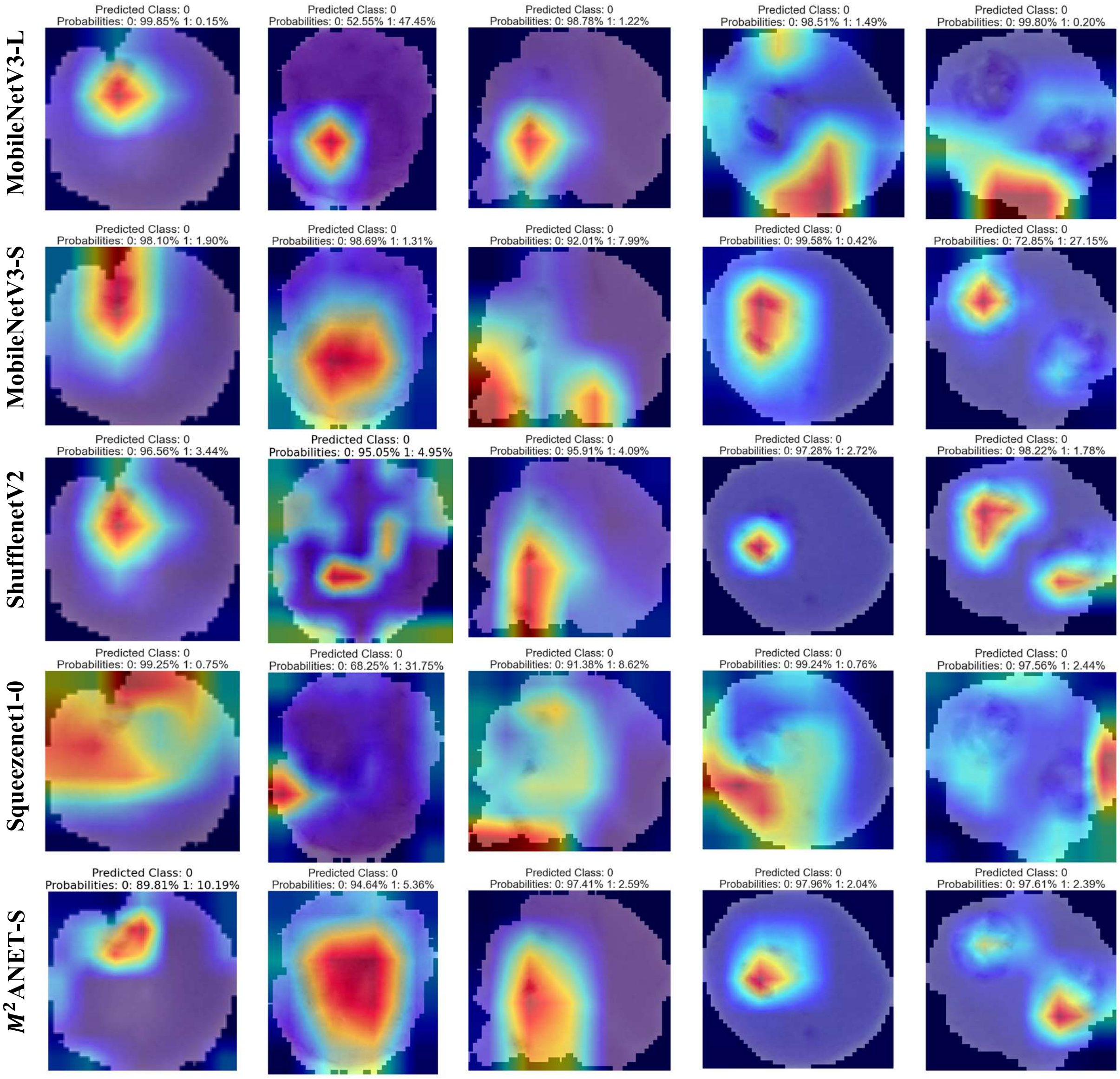

MobileNetV3-L
Predicted Class: 0
Probabilities: 0: 99.85% 1: 0.15%
Predicted Class: 0
Probabilities: 0: 52.55% 1: 47.45%
Predicted Class: 0
Probabilities: 0: 98.78% 1: 1.22%
Predicted Class: 0
Probabilities: 0: 98.51% 1: 1.49%
Predicted Class: 0
Probabilities: 0: 99.80% 1: 0.20%
MobileNetV3-S
Predicted Class: 0
Probabilities: 0: 98.10% 1: 1.90%
Predicted Class: 0
Probabilities: 0: 98.69% 1: 1.31%
Predicted Class: 0
Probabilities: 0: 92.01% 1: 7.99%
Predicted Class: 0
Probabilities: 0: 99.58% 1: 0.42%
Predicted Class: 0
Probabilities: 0: 72.85% 1: 27.15%
ShufflenetV2
Predicted Class: 0
Probabilities: 0: 96.56% 1: 3.44%
Predicted Class: 0
Probabilities: 0: 95.05% 1: 4.95%
Predicted Class: 0
Probabilities: 0: 95.91% 1: 4.09%
Predicted Class: 0
Probabilities: 0: 97.28% 1: 2.72%
Predicted Class: 0
Probabilities: 0: 98.22% 1: 1.78%
Squeezenet1-0
Predicted Class: 0
Probabilities: 0: 99.25% 1: 0.75%
Predicted Class: 0
Probabilities: 0: 68.25% 1: 31.75%
Predicted Class: 0
Probabilities: 0: 91.38% 1: 8.62%
Predicted Class: 0
Probabilities: 0: 99.24% 1: 0.76%
Predicted Class: 0
Probabilities: 0: 97.56% 1: 2.44%
M^2ANET-S
Predicted Class: 0
Probabilities: 0: 89.81% 1: 10.19%
Predicted Class: 0
Probabilities: 0: 94.64% 1: 5.36%
Predicted Class: 0
Probabilities: 0: 97.41% 1: 2.59%
Predicted Class: 0
Probabilities: 0: 97.96% 1: 2.04%
Predicted Class: 0
Probabilities: 0: 97.61% 1: 2.39%

Figure 3 Comparing GRAD-CAM visualization with several methods

ROC & Precision-Recall Curve. M^2ANET achieves an ROC value of 0.95 in Figure 4, the highest among compared methods. This indicates its effectiveness in distinguishing infected cells from uninfected ones by maintaining a high true positive rate while minimizing false positives. Similarly, achieving a precision-recall curve score of 0.96 in Figure 5 provides valuable insight into the trade-off between precision and recall, further highlighting the model's performance in identifying infected cells.

Computational complexity. M^2ANET-S and L variants, shows promising results in computational complexity compared to SOTA methods architectures. M^2ANET-L achieves competitive performance with relatively lower parameter count and file size compared to models like ResNet-18 and MobileNetV3-L, while maintaining efficient latency and throughput. Similarly, M^2ANET-S is lighter with significantly reduced latency, making it suitable for real-time applications without compromising on model accuracy. These results in Table 1 show the potential of M^2ANET as efficient alternatives for practical deployment in resource-constrained environments.

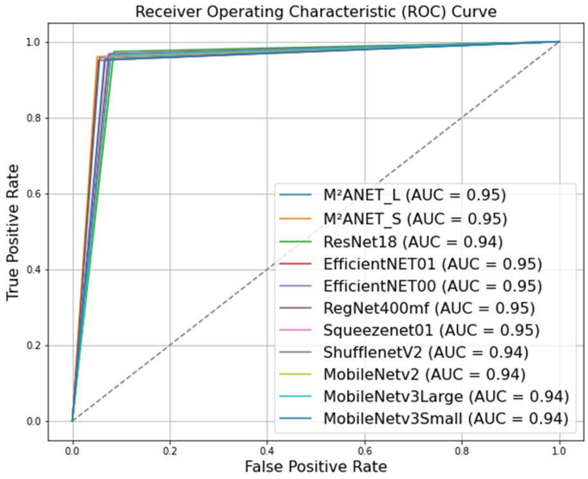


Figure 4 ROC

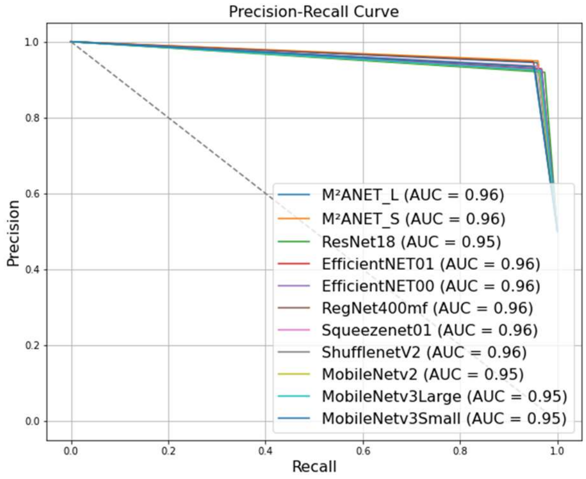


Figure 5 Precision-Recall Curve

Table 1 Computational complexity

Model	#Params	#FLOPs	File size	Latency	Throughput img/sec
ResNet-18	11.2m	0.49	**3.8 mb**	**0.005s**	**10401**
Efficient-B0	4.0m	**0.11**	16.0 mb	0.018s	2421
Efficent-B1	6.5m	0.17	25.9 mb	0.027s	2357
RegNet400mf	3.9m	**0.11**	15.5 mb	0.021s	1603
$\boldsymbol{M^2}$**ANET-L**	**2.2m**	2.73	9.9 mb	0.009s	7157
MobileNetV2	2.2m	0.09	8.9 mb	0.012s	5551
MobileNetV3-L	4.2m	0.06	16.7 mb	0.022s	2550
MobileNetV3-S	1.5m	**0.02**	6.1 mb	0.012s	4436
ShufflenetV2	1.2m	0.04	5.1 mb	0.014s	3984
Squeezenet1-0	**0.7m**	0.17	**2.9 mb**	0.004s	**15557**
$\boldsymbol{M^2}$**ANET-S**	1.2m	2.42	6.1 mb	**0.0015s**	8023

Classification accuracy and comparison. Table 2 presents the top-1 accuracy and Cohen Kappa scores of various models including M^2ANET, showing the performance in classification tasks. Overall, M^2ANET-S achieved the highest top-1 accuracy of 95.45% and a Cohen Kappa score of 0.91, indicating its superior performance in accurately classifying parasitized cell images and non-parasitized images. Notably, ResNet-18, Efficient-B0, Efficient-B1, and Squeezenet1-0 also demonstrate strong performance, with top-1 accuracy scores ranging from 94.42% to 94.86% and Cohen Kappa scores around 0.89 to 0.90. These findings highlight the effectiveness of M^2ANET model, in achieving high accuracy and reliability in disease classification tasks of Plasmodium in cell images compared to some of the SOTA architectures such as ResNet, EfficientNet, and MobileNet variants.

Sensitivity and Specificity. Table 3 shows the performance of various models evaluated using a 5-fold cross-validation approach, with True Positive Rate (TPR) and True Negative Rate (TNR) as the key metrics. TPR is a measure of sensitivity, indicating the model's ability to correctly identify positive cases, while TNR is a measure of specificity, indicating the model's ability to correctly identify negative cases.

Table 2 top-1 accuracy and Cohen Kappa

	Model	Input	Epoch	F1-score	Recall	Precision	Top-1 acc.	Cohen Kappa
Lightweight	ResNet-18	112	90	**0.95**	**0.97**	0.92	94.42%	0.89
	Efficient-B0	112	90	**0.95**	**0.97**	0.93	94.57%	0.89
	Efficent-B1	112	90	**0.95**	**0.97**	0.93	94.64%	0.89
	RegNet400mf	112	90	**0.95**	0.95	**0.95**	94.86%	**0.90**
	M^2ANET-L	112	90	**0.95**	0.96	0.94	**95.11%**	**0.90**
Mobile based	MobileNetV2	112	90	0.94	**0.96**	0.92	94.10%	0.88
	MobileNetV3-L	112	90	0.94	**0.96**	0.92	94.27%	0.89
	MobileNetV3-S	112	90	0.94	0.95	0.93	94.22%	0.88
	ShufflenetV2	112	90	0.94	0.95	0.93	94.36%	0.89
	Squeezenet1-0	112	90	**0.95**	**0.96**	0.94	94.70%	0.89
	M^2ANET-S	112	90	**0.95**	**0.96**	**0.95**	**95.45%**	**0.91**

Comparative analysis of models. ResNet-18 demonstrates high sensitivity, ranging from 96.88% to 97.77%, indicating strong performance in identifying positive cases. However, its specificity ranges from 90.90% to 92.09%, which, although consistent, is lower compared to other models like M^2ANET-S, indicating less accuracy in identifying negative cases.

- EfficientNet-B0 shows sensitivity ranging from 96.17% to 97.34%, showing high accuracy in positive case identification, similar to ResNet-18. Specificity for EfficientNet-B0 ranges from 91.88% to 92.89%, which is slightly higher than ResNet-18, demonstrating better performance in negative case identification.

- EfficientNet-B1 achieved a high consistent sensitivity, ranging from 95.95% to 97.20%, EfficientNet-B1 matches the positive case identification capabilities of the previous models. Its specificity, ranging from 91.98% to 93.26%, indicates an improvement over both ResNet-18 and EfficientNet-B0 in negative case identification.
- RegNet-400MF shows lower sensitivity compared to other models, ranging from 93.83% to 95.66%. However, its specificity, ranging from 94.28% to 95.31%, is among the highest, making it excellent at identifying negative cases but less effective at identifying positive cases compared to others.
- M^2ANET-L achieved sensitivity from 95.32% to 96.48%, and specificity from 93.91% to 94.51%. M^2ANET-L balances well between high positive case identification and good negative case identification, making it a robust model for both metrics.
- The sensitivity for MobileNetV2 ranges from 95.74% to 96.84%, and specificity ranges from 90.84% to 92.82%. It performs similarly to ResNet-18 in positive case identification but has a slightly lower specificity, indicating less accuracy in identifying negative cases.
- MobileNetV3-L sensitivity ranges from 95.95% to 96.79%, with specificity from 91.53% to 92.53%. While it performs well in positive case identification, it does not outperform M^2ANET-L or EfficientNet-B1 in negative case identification.
- Sensitivity for MobileNetV3-S ranges from 94.68% to 95.74%, slightly lower than the L variant. Specificity ranges from 92.87% to 93.70%, indicating stable but not outstanding performance in negative case identification compared to larger models.
- ShuffleNetV2 shows variability in sensitivity, ranging from 94.11% to 96.12%, and specificity from 92.87% to 94.14%. ShuffleNetV2 is consistent in negative case identification but shows some variation in identifying positive cases.
- SqueezeNet1-0 has a sensitivity score ranging from 95.52% to 96.55%, and specificity from 92.87% to 93.85%, SqueezeNet1-0 performs similarly to MobileNetV3-L, indicating high accuracy in positive case identification and stable performance in negative case identification.

M^2ANET-S achieves better sensitivity ranging from 95.75% to 96.42%, and specificity from 93.99% to 96.05%. The model not only achieves high accuracy in identifying positive cases but also shows the highest performance in negative case identification among all models, making it the most balanced and reliable model for both metrics.

Table 3 Sensitivity and Specificity using 5-fold cross validation

	Model	k-fold 1		k-fold 2		k-fold 3		k-fold 4		k-fold 5	
		TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR	TPR	TNR
Lightweight	ResNet-18	**96.88**	91.46	**97.34**	91.58	**97.77**	92.09	**97.50**	90.90	**97.46**	91.25
	Efficient-B0	96.17	92.58	96.48	92.01	97.34	92.89	97.21	92.33	96.86	91.88
	Efficent-B1	95.95	91.98	96.84	93.26	97.20	92.67	97.13	92.54	96.64	92.24
	RegNet400mf	93.83	**94.28**	95.40	**95.31**	95.18	**94.73**	95.66	94.41	95.37	**94.50**
	M^2ANET-L	95.32	93.91	95.76	94.51	96.48	94.51	95.81	**94.48**	95.89	94.42
Mobile based	MobileNetV2	95.81	91.83	96.33	90.84	**96.84**	92.82	96.32	92.19	95.74	92.24
	MobileNetV3-L	95.95	91.91	**96.62**	92.23	96.33	92.53	**96.76**	92.11	**96.79**	91.53
	MobileNetV3-S	94.68	92.87	95.11	93.48	95.26	93.70	95.74	93.12	95.14	93.08
	ShufflenetV2	94.11	92.87	96.12	93.11	95.97	**94.14**	95.22	93.19	95.74	93.15
	Squeezenet1-0	95.53	92.87	96.55	93.85	96.48	93.63	96.03	93.19	95.52	93.37
	M^2ANET-S	**96.07**	**94.50**	95.80	**94.71**	96.07	93.99	95.75	**96.05**	96.42	**95.15**

Sensitivity and specificity are key metrics for evaluating the performance of medical diagnostic models. High sensitivity is crucial for detecting as many positive cases as possible, thereby reducing the risk of missed diagnoses. High specificity ensures that negative cases are correctly identified, preventing unnecessary anxiety and treatment. A balanced approach between these metrics is essential for creating reliable and efficient diagnostic tools, particularly in resource-contained settings, like mobile and edge devices. Understanding and optimizing these metrics can significantly enhance the effectiveness of medical diagnostic systems like M^2ANET in identifying conditions such as plasmodium parasitized cells.

4.2 Ablation study on the effect of each component

Table 4 Ablation study on the effect of each component

Settings	Component	Layers	#Params	FLOPs	Accuracy
(a)	MBconv3	[8]	8.5m	3.05G	92.76
(b)	MBconv3 + MHSA	[4, 4]	2.2m	2.73G	95.11
(d)	MBconv3 + MHSA	[4, 2]	**1.2m**	**2.42G**	**95.45**

Using only MBConv3 layers resulted in an accuracy of 92.76%, with 8.5 million parameters and 3.05 GFLOPs. Integrating 4 MBConv3 layers with 4 MHSA layers improved accuracy to 95.11%, while reducing parameters to 2.2 million and FLOPs to 2.73 GFLOPs. Further reducing MHSA layers to 2 while maintaining 4 MBConv3 layers achieved the highest accuracy of 95.45%, with only 1.2 million parameters and 2.42 GFLOPs, indicating optimal efficiency and performance.

5 Conclusion

This work introduces M^2ANET, a novel mobile hybrid model for classifying Plasmodium parasites in infected cell images. By integrating convolutional layers and attention mechanisms, the model achieves a balance between high classification accuracy and computational efficiency, making it suitable for resource-constrained settings such as mobile and edge devices. Its effectiveness in identifying infected cell images positions it as a promising tool for improving malaria diagnosis. Future work should focus on validating its applicability in real-world clinical settings and exploring its scalability for large-scale deployment.

Limitations. Firstly, the models were trained and tested on a single malaria blood smear dataset, which limits their generalizability to other datasets or real-world conditions. The hybrid design of convolutional layers and self-attention mechanisms, while effective, may reduce the interpretability of the model's decisions, which is critical in medical applications. Furthermore, the study focuses solely on malaria detection, limiting its applicability to other medical image classification tasks without extensive retraining. Finally, the performance of the models in real-world diagnostic pipelines, including data preprocessing and system integration, remains unexplored. Future research should address these limitations to enhance the models' robustness and generalization.

Disclosure of Interests. The authors declare no competing interests.

Funding. This research received no external funding.

References

[1] P.E. Hart, N.J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans. System Science and Cybernetics SSC-4*, 2:100-107, 1968.

[2] Richard E. Korf. Iterative-deepening A*: an optimal admissible tree search. In *Proc. of the IX Int. Joint Conf.. on Artificial Intelligence (IJCAI'85)*, pages 1034-1036, 1985.

[3] Tobias Oetiker, Hubert Partl, Irene Hyna, and Elisabeth Schlegl. *The not so short introduction to LaTeX2e,* 2008.

[4] Judea Pearl. *Heuristics.* Addison-Wesley, Readin, Massachusetts, 1984.

[5] V. Vovk. Competing with wild prediction rules, *Machine Learning*, 69:193--212, 2007. doi: 10.1007/s10994-007-5021-y

[1] Skorokhod, Oleksii, Ekaterina Vostokova, and Gianfranco Gilardi. "The role of P450 enzymes in malaria and other vector-borne infectious diseases." BioFactors (2023).

[2] Baptista, Vitoria, Mariana S. Costa, Carla Calcada, Miguel Silva, Jose Pedro Gil, Maria Isabel Veiga, and Susana O. Catarino. "The future in sensing technologies for malaria surveillance: a review of hemozoin-based diagnosis." ACS sensors 6, no. 11 (2021): 3898-3911.

[3] Matthews, Jerrid, Rajan Kulkarni, Mario Gerla, and Tammara Massey. "Rapid dengue and outbreak detection with mobile systems and social networks." Mobile Networks and Applications 17 (2012): 178-191.

[4] Maduako, Chidinma. "Malaria diagnosis based on a machine learning system." (2020).

[5] Rashmi, R. "A Comparative Study of Blood Smear, Quantitative Buffy Coat and Antigen Detection for Diagnosis of Malaria." PhD diss., Rajiv Gandhi University of Health Sciences (India), 2013.

[6] Poostchi, Mahdieh, Kamolrat Silamut, Richard J. Maude, Stefan Jaeger, and George Thoma. "Image analysis and machine learning for detecting malaria." Translational Research 194 (2018): 36-55.

[7] Ghosh, Pramit, Debotosh Bhattacharjee, and Mita Nasipuri. "Automatic system for plasmodium species identification from microscopic images of blood-smear samples." Journal of Healthcare Informatics Research 1 (2017): 231-259.

[8] Grochowski, Michał, Michał Wąsowicz, Agnieszka Mikołajczyk, Mateusz Ficek, Marek Kulka, Maciej S. Wróbel, and Małgorzata Jędrzejewska-Szczerska. "Machine learning system for automated blood smear analysis." Metrology and Measurement Systems 26, no. 1 (2019).

[9] Esteva, Andre, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher. "Deep learning-enabled medical computer vision." NPJ digital medicine 4, no. 1 (2021): 5.

[10] Wang, Zhiqiong, Yiqi Luo, Junchang Xin, Hao Zhang, Luxuan Qu, Zhongyang Wang, Yudong Yao, Wancheng Zhu, and Xingwei Wang. "Computer-aided diagnosis based on extreme learning machine: a review." IEEE Access 8 (2020): 141657-141673.

[11] Rajaraman, Sivaramakrishnan, Sameer K. Antani, Mahdieh Poostchi, Kamolrat Silamut, Md A. Hossain, Richard J. Maude, Stefan Jaeger, and George R. Thoma. "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images." PeerJ 6 (2018): e4568.

[12] Elangovan, Poonguzhali, and Malaya Kumar Nath. "A novel shallow convnet-18 for malaria parasite detection in thin blood smear images: Cnn based malaria parasite detection." SN computer science 2, no. 5 (2021): 380.

[13] Nakasi, Rose, Ernest Mwebaze, Aminah Zawedde, Jeremy Tusubira, Benjamin Akera, and Gilbert Maiga. "A new approach for microscopic diagnosis of malaria parasites in thick blood smears using pre-trained deep learning models." SN Applied Sciences 2 (2020): 1-7.

[14] Diyasa, I. Gede Susrama Mas, Akhmad Fauzi, Ariyono Setiawan, Moch Idhom, Radical Rakhman Wahid and Alfath Daryl Alhajir. "Pre-trained deep convolutional neural network for detecting malaria on the human blood

smear images." In 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 235-240. IEEE, 2021.

[15] Swastika, W., G. M. Kristianti, and R. B. Widodo. "Effective preprocessed thin blood smear images to improve malaria parasite detection using deep learning." In Journal of Physics: Conference Series, vol. 1869, no. 1, p. 012092. IOP Publishing, 2021.

[16] Araujo, F. A. S., N. D. Colares, U. P. Carvalho, C. F. F. Costa Filho, and M. G. F. Costa. "Plasmodium Life Cycle-Stage Classification on Thick Blood Smear Microscopy Images using Deep Learning: A Contribution to Malaria Diagnosis." In 2023 19th International Symposium on Medical Information Processing and Analysis (SIPAIM), pp. 1-4. IEEE, 2023.

[17] Widodo, Sri. "Texture analysis to detect malaria tropica in blood smears image using support vector machine." International Journal of Innovative Research in Advanced Engineering 1, no. 8 (2014): 301-306.

[18] Alharbi, Amal H., Meng Lin, B. Ashwini, Mohamed Yaseen Jabarulla, and Mohd Asif Shah. "Detection of peripheral malarial parasites in blood smears using deep learning models." Computational Intelligence and Neuroscience 2022 (2022).

[19] Shetty, Vijaya, and Vijaylaxmi Kochari. "Detection and classification of peripheral plasmodium parasites in blood smears using filters and machine learning algorithms." Proceedings http://ceur-ws. org ISSN 1613 (2023): 0073.

[20] Jia, Jianguo, Wen Liang, and Youzhi Liang. "A review of hybrid and ensemble in deep learning for natural language processing." arXiv preprint arXiv:2312.05589 (2023).

[21] Yunusa, Haruna, Shiyin Qin, Abdulrahman Hamman Adama Chukkol, Abdulganiyu Abdu Yusuf, Isah Bello, and Adamu Lawan. "Exploring the Synergies of Hybrid CNNs and ViTs Architectures for Computer Vision: A survey." arXiv preprint arXiv:2402.02941 (2024).

[22] Yang, Jianwei, Chunyuan Li, Pengchuan Zhang, Xiyang Dai, Bin Xiao, Lu Yuan, and Jianfeng Gao. "Focal self-attention for local-global interactions in vision transformers." arXiv preprint arXiv:2107.00641 (2021).

[23] Howard, Andrew, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang et al. "Searching for mobilenetv3." In Proceedings of the IEEE/CVF international conference on computer vision, pp. 1314-1324. 2019.

[24] Ma, Ningning, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. "Shufflenet v2: Practical guidelines for efficient cnn architecture design." In Proceedings of the European conference on computer vision (ECCV), pp. 116-131. 2018.

[25] Iandola, Forrest N., Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." arXiv preprint arXiv:1602.07360 (2016).

[26] Bibin, Dhanya, Madhu S. Nair, and P. Punitha. "Malaria parasite detection from peripheral blood smear images using deep belief networks." IEEE Access 5 (2017): 9099-9108.

[27] Sivaramakrishnan, Rajaraman, Sameer Antani, and Stefan Jaeger. "Visualizing deep learning activations for improved malaria cell classification." In Medical informatics and healthcare, pp. 40-47. PMLR, 2017.

[28] Rajaraman, Sivaramakrishnan, Sameer K. Antani, Mahdieh Poostchi, Kamolrat Silamut, Md A. Hossain, Richard J. Maude, Stefan Jaeger, and George R. Thoma. "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images." PeerJ 6 (2018): e4568.

[29] Yang, Feng, Mahdieh Poostchi, Hang Yu, Zhou Zhou, Kamolrat Silamut, Jian Yu, Richard J. Maude, Stefan Jaeger, and Sameer Antani. "Deep learning for smartphone-based malaria parasite detection in thick blood smears." IEEE journal of biomedical and health informatics 24, no. 5 (2020): 1427-1438.

[30] Vijayalakshmi, A. "Deep learning approach to detect malaria from microscopic images." Multimedia Tools and Applications 79, no. 21 (2020): 15297-15317.

[31] Zedda, Luca, Andrea Loddo, and Cecilia Di Ruberto. "A deep learning based framework for malaria diagnosis on high variation data set." In International Conference on Image Analysis and Processing, pp. 358-370. Cham: Springer International Publishing, 2022.

[32] Madhu, Golla, A. Govardhan, Vinayakumar Ravi, Sandeep Kautish, B. Sunil Srinivas, Tanupriya Chaudhary, and Manoj Kumar. "DSCN-net: a deep Siamese capsule neural network model for automatic diagnosis of malaria parasites detection." Multimedia Tools and Applications 81, no. 23 (2022): 34105-34127.

[33] Siłka, Wojciech, Michał Wieczorek, Jakub Siłka, and Marcin Woźniak. "Malaria detection using advanced deep learning architecture." Sensors 23, no. 3 (2023): 1501.

[34] Abdurahman, Fetulhak, Kinde Anlay Fante, and Mohammed Aliy. "Malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models." BMC bioinformatics 22 (2021): 1-17.

[35] Zhong, Yuming, Ying Dan, Yin Cai, Jiamin Lin, Xiaoyao Huang, Omnia Mahmoud, Eric S. Hald, Akshay Kumar, Qiang Fang, and Seedahmed S. Mahmoud. "Efficient Malaria Parasite Detection From Diverse Images of Thick Blood Smears for Cross-Regional Model Accuracy." IEEE Open Journal of Engineering in Medicine and Biology (2023).

[36] Aris, Thaqifah Ahmad, Aimi Salihah Abdul Nasir, Lim Chee Chin, H. Jaafar, and Z. Mohamed. "Fast k-means clustering algorithm for malaria detection in thick blood smear." In 2020 IEEE 10th International Conference on System Engineering and Technology (ICSET), pp. 267-272. IEEE, 2020.

[37] Jahan, Rashke, and Shahzad Alam. "Improving Classification Accuracy Using Hybrid Machine Learning Algorithms on Malaria Dataset." Engineering Proceedings 56, no. 1 (2023): 232.

[38] Murmu, Anita, and Piyush Kumar. "DLRFNet: deep learning with random forest network for classification and detection of malaria parasite in blood smear." Multimedia Tools and Applications (2024): 1-23.

[39] Rajaraman, Sivaramakrishnan, Sameer K. Antani, Mahdieh Poostchi, Kamolrat Silamut, Md A. Hossain, Richard J. Maude, Stefan Jaeger, and George R. Thoma. "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images." PeerJ 6 (2018): e4568.

[40] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[41] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." In International conference on machine learning, pp. 6105-6114. PMLR, 2019.

[42] Radosavovic, Ilija, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollár. "Designing network design spaces." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10428-10436. 2020.

[43] Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-cam: Visual explanations from deep networks via gradient-based localization." In Proceedings of the IEEE international conference on computer vision, pp. 618-626. 2017.