I. J. Engineering and Manufacturing, **2024, 6, 11-26**
Published Online on December 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2024.06.02

Dilated Convolutional Neural Network with Attention Mechanism for Classification of Malaria Parasites

Suleiman Garba*
Department of Computer Science, Federal Capital Territory College of Education Zuba-Abuja, Nigeria
Email: sulgarba@gmail.com
ORCID iD: https://orcid.org/0000-0001-7937-2816
*Corresponding author

Muhammad Bashir Abdullahi
Department of Computer Science, Federal University of Technology, Minna, Nigeria
Email: el.bashir02@futminna.edu.ng
ORCID iD: https://orcid.org/0000-0001-8200-2787

Sulaimon Adebayo Bashir
Department of Computer Science, Federal University of Technology, Minna, Nigeria
Email: bashirsulaimon@futminna.edu.ng
ORCID iD: https://orcid.org/0000-0001-8690-3953

Abisoye Opeyemi Aderike
Department of Computer Science, Federal University of Technology, Minna, Nigeria
Email: o.abisoye@futmuinna.edu.ng
ORCID iD: https://orcid.org/0000-0002-7324-2479

Received: 16 February, 2024; Revised: 21 May, 2024; Accepted: 23 October, 2024; Published: 08 December, 2024

Abstract: Malaria remains a pervasive global health challenge, affecting millions of lives daily. Traditional diagnostic methods, involving manual blood smear examination, are time-consuming and prone to errors, especially in large-scale testing. Although promising, automated detection techniques often fail to capture the intricate spatial features of malaria parasites leading to inconsistent performance. In order to close these gaps, this work suggest an improved technique that combines a Self-Attention Mechanism and a Dilated Convolutional Neural Network (D-CNN) to allow the model to effectively and precisely classify malaria parasites as infected or uninfected. Both local and global spatial information are captured by dilated convolutions, and crucial features are given priority by the attention mechanism for accurate detection in complex images. We also examine batch size variation and find that it plays a crucial role in maximizing generalization, accuracy, and resource efficiency. A batch size of 64 produced superior results after testing six different sizes, yielding an AUC of 99.12%, F1-Score of 96, precision of 97.63%, recall of 93.99%, and accuracy of 96.08%. This batch size balances efficient gradient updates and stabilization, reducing overfitting and improving generalization, especially on complex medical datasets. Our approach was benchmarked against existing competitors using the same publicly available malaria dataset, demonstrating a 2-3% improvement in AUC and precision over state-of-the-art models, such as traditional CNNs and machine learning methods. This highlights its superior ability to minimize false positives and negatives, particularly in complex diagnostic cases. These advancements enhance the reliability of large-scale diagnostic systems, improve clinical decision-making, and address key challenges in automated malaria detection.

Index Terms: Attention Mechanism, Batch Size, Classification, Dilated CNN, Malaria, Parasite.

1. Introduction

Parasitic diseases represent a significant threat to global public health, particularly in regions like Africa, where infectious and parasitic diseases rank among the leading causes of death. According to the World Health Organization's (WHO) Global Health Estimates in 2020 [1], there were over 200 million reported cases globally and more than 400,000 documented deaths attributed to these diseases. Among them, malaria, caused by the Plasmodium parasite transmitted through the bite of infected Anopheles mosquitoes, stands out as a particularly fatal illness, destroying red blood cells and leading to severe health complications [2].

The primary method for diagnosing malaria infection is the peripheral blood smear (PBS) test, a cost-effective approach involving the preparation of a thin layer of red blood cells from a patient's blood sample on a glass slide. These slides are then stained with dyes like Leishman or Giemsa and examined under a microscope to identify parasites and anomalies within the red blood cells [3,4]. Microscopic analysis provides crucial information about the presence of infection, the specific Plasmodium species involved, and the progression of the infection.

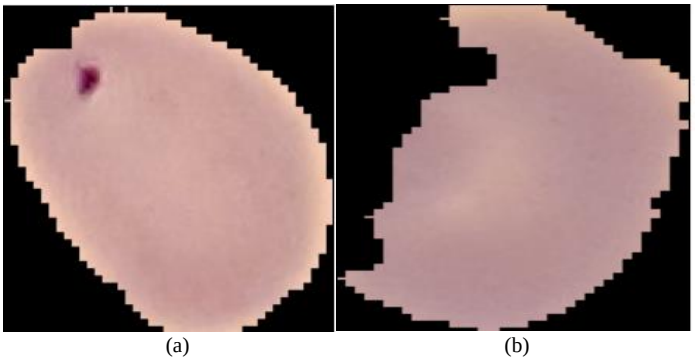
Recent advancements have shown that Deep Learning (DL) techniques, particularly Convolutional Neural Networks (CNNs), offer highly effective means for classifying medical images[5], eliminating the need for manual feature engineering. CNNs excel in capturing local features of images [6], leveraging spatial and hierarchical structures to extract pertinent patterns. However, their ability to capture global features, especially in images with sparse areas, is limited.

Moreover, existing CNN-based models often lack flexibility in simultaneously extracting both local and global features [7]. To address these limitations and enhance the classification of malaria parasites, this study proposes an integrated model combining Dilated CNN and attention mechanism. Dilated CNN extends the receptive field of convolution outputs to capture a broader range of information, overcoming the constraints of traditional CNNs with small convolutional kernels. Meanwhile, attention mechanisms localize and refine complex image features relevant to the classification task. The fusion of these techniques allows for the extraction of spatial features by the Dilated CNN, refined by the attention mechanism, and further classified using fully connected layers.

In view of these, dilated convolutions efficiently expand the receptive field, and self-attention focuses on important areas for accurate parasite detection, making this combination of Dilated Convolutional Neural Networks (D-CNNs) and a Self-Attention Mechanism ideal for diagnosing malaria. D-CNNs handle smaller, medical-specific datasets more effectively than transformers, which need large amounts of data and resources.

The key contributions of this work include:

(1) Proposing a novel framework integrating Dilated CNN and attention mechanism for malaria parasite classification.
(2) Enhancing classification performance through optimization of batch size hyperparameters for distinguishing between infected and uninfected malaria parasite cell images.
(3) Validating the proposed approach through comparative analysis with existing methodologies, demonstrating its viability and effectiveness in malaria diagnosis.



(a) (b)

Fig. 1. (a) Infected Image; (b) Uninfected Image

The paper is structured as follows: Section 1 provides the introduction, Section 2 presents the literature review, Section 3 outlines the methodology employed for the proposed model, Section 4 details the experimental results and analysis, and Section 5 encompasses the concluding remarks.

2. Related Work

Recent studies have explored the automated recognition and categorization of red blood cells with and without malaria infection utilizing digital pictures of blood smears. Feature extraction, picture segmentation, and classification are important procedures in these investigations. Purnama et al. [8] used hue channel shifting, the HSV histogram, and genetic programming-based classification on a comparatively small dataset of 180 samples. Although the approach achieved a 95.49% correct identification rate for parasites and 95.58% correct identification rate for non-parasites, its heavy reliance on manual feature extraction may cause it to miss subtle but significant features. A deep network was trained on 4,100 samples by Bibin et al. [9], yielding an F-score of 89.66, a sensitivity of 97.60%, and a specificity of 95.92%. Nevertheless, the model's ability to be applied to bigger, more varied datasets is constrained by the small dataset size.

In order to solve this problem, Rajaraman et al. [10] extracted features from 27,558 samples, identifying Plasmodium from red blood cell (RBC) segmented images with a 92.7% accuracy rate. Their method highlights the benefit of larger datasets by exhibiting superior generalization. With the help of cutting-edge deep learning models like Xception, Inception-V3, ResNet-50, VGG-16, and AlexNet, Sriporn et al. [11] were able to achieve 99.28% accuracy. However, when trained on small datasets, deep models like these are prone to overfitting, which limits their applicability. Moreover, clinical adoption may be impeded by the limited explainability of these models, as noted in the research of Rajaraman et al. [10] and Islam et al. [15].

Goni et al. [12] and Khan et al. [13] showed the superiority of deep learning models over conventional methods in automated feature extraction. Goni et al. [12] obtained accuracy rates of 96.79% and 99.56% for the original and modified datasets, respectively, using a CNN with a double hidden layer ELM approach. Recall rates of 86%, 95%, and 91.66% were reported by Khan et al. [13] using neural network (CNN), support vector machine (SVM), and random forest (RF) models. Notwithstanding these findings, the interpretability issue with deep learning models persists in the field of medical diagnostics.

Fuhad et al. [14] demonstrated high precision and recall by using RBC images to diagnose malaria with 99.23% accuracy using k-nearest neighbors (KNNs) and SVM. But when employing smaller datasets, Mohanty et al. [16] emphasized the danger of overfitting with intricate models like autoencoder (AE) and self-organizing map (SOM). Similar to this, Dong et al. [17] used transfer learning (TL) models like GoogLeNet and AlexNet and achieved a 95% accuracy rate; however, the use of pre-trained models that were created using non-medical datasets may not adequately capture the subtleties unique to malaria.

Li et al. [19] improve spatial feature detection and attain higher accuracy in difficult images by introducing a Residual Attention learning network coupled with SVM (RAL-CNNSVM). This was further developed by Islam et al. [15] using a multiheaded attention-based transformer model that performed better than traditional CNNs by concentrating on important regions of the image and efficiently handling large amounts of data. While the practical application in resource-constrained environments remains a challenge, Reddy et al. [21] and Gummadi, Ghoosh, and Vootla [26] applied transfer-learning-based CNN architectures, showing promise for faster training on medical datasets. In order to improve model robustness and lower noise, Maqsood et al. [24], Shal & Gupta [25] proposed augmented learning techniques like image augmentation and bilateral filtering, which resulted in accuracy rates of 94.67% and 96.82%, respectively. These studies highlight the need for stronger feature extraction and reliable generalization in malaria detection systems.

Further improvements are still required, even though the previously mentioned literature highlights the efficacy of deep learning models in malaria detection, achieving high accuracy with various methodologies. Despite their widespread use, traditional CNNs frequently suffer from poor spatial awareness and may overlook important details in medical images. By enabling models to concentrate on the most important regions of the image, attention mechanisms can help with these problems by enhancing the interpretability and detection of spatial features. To create more durable and dependable diagnostic systems for real-world applications, issues like overfitting, reliance on small datasets, limited explainability, and transfer learning limitations still need to be resolved.

3. Materials and Methods

The model architecture, illustrated in Fig. 2, comprises three dilated layers for feature extraction, followed by max pooling. The resulting features are then flattened into one-dimensional feature vectors. Subsequently, an attention block processes the output of the dilation convolution block. The output of the attention block is fed into dense layers with a dropout rate of 0.2 to mitigate overfitting and facilitate classification into infected and uninfected categories.

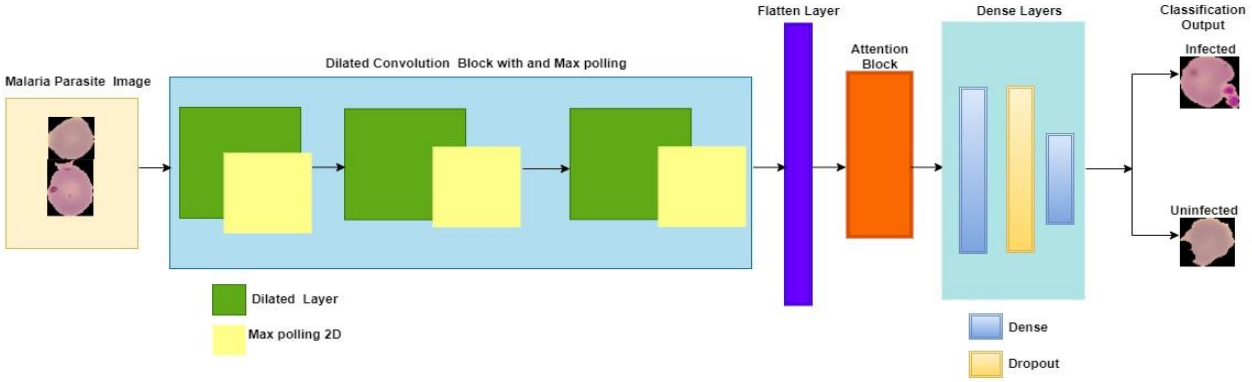


Fig. 2. Proposed Dilated CNN-Attention Mechanism Architecture

3.1 Data set descriptions

Malaria images for this study were obtained from a publicly accessible dataset provided by the USA National Institutes of Health (NIH), chosen for its suitability in facilitating comparative analyses with related work, thereby establishing a standard. The dataset comprised a total of 27,699 images (as shown in Table 1), consisting of 14,394 infected and 13,305 uninfected samples.

To prepare the data for training and testing, normalization procedures were applied. The images were resized to dimensions of 150 by 150 pixels, ensuring uniformity in width and height. All images maintained three color channels throughout the dataset. Subsequently, the dataset was divided into training and testing sets for further analysis.

Table 1. Data size for Infected and uninfected Parasites

Infected	14394
Uninfected	13305
Total	**27699**

3.2 Image pre-processing

Grayscale images were used to simplify the model and focus on the structural features of malaria parasites, which are crucial for accurate classification. By reducing the input dimensionality from three channels (RGB) to one, the model complexity and computational load are minimized, helping to avoid overfitting, especially with limited datasets. Additionally, since the distinguishing features of malaria parasites, such as shape, size, and texture, are adequately represented in grayscale, the potential benefits of color information are outweighed by the advantages of a more streamlined and efficient model. This approach enables the enhanced dilated CNN and self-attention mechanism to effectively capture and analyze the relevant morphological features without being distracted by color variations, which are less significant for this specific classification task.

Oversampling techniques SMOTE was used for the minority class and apply class weighting in the loss function during model training to enhance balance because malaria datasets frequently have significant class imbalances between infected and uninfected samples, which can cause bias model performance.

To enhance the input image quality for subsequent processing tasks, Python 3.7 functions were employed. The image underwent a conversion from color to grayscale, aimed at eliminating or minimizing any existing noise. The resultant image after the pre-processing steps is depicted in Fig. 3.

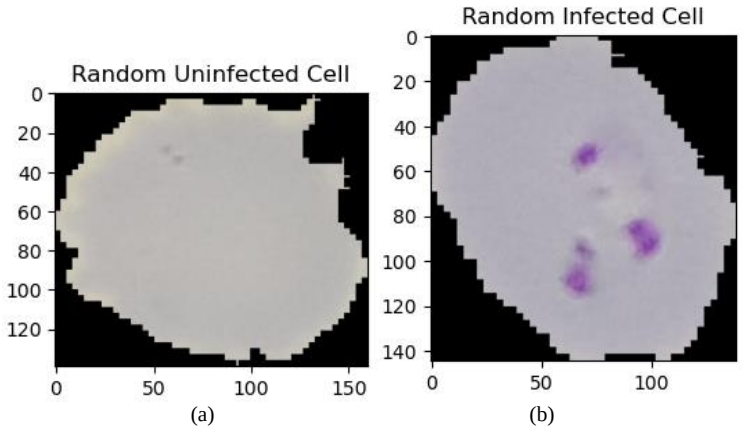


Fig. 3. (a) Pre processed Uninfected Image; (b) Pre processed Infected Image

3.3 Proposed Model

Recognizing the importance of accurately categorizing malaria-infected blood cells, there is a need for precise and practical methods to support diagnosis and long-term treatment solutions. The incorporation of attention mechanism into this model was due to its widespread implementation in some relevant literatures. This work referenced [27, 28, 29], as depicted in Fig. [4, 5, 6], have inspired the development of this proposed model, aiming to address their identified limitations, such as the need for further performance evaluation [28 and 29], as well as issues related to the loss of fine-grained information and spatial resolution [27].

The proposed model consists of two main components: the dilated convolution block and the attention block. The dilated convolution block, depicted in Fig. 2, comprises of three dilated convolution layers with max pooling. The first dilated block (D1) receives the input image (x) which is subsequently pass on to the next block (D2) and then to (D3). The last layer output is then flatten and then pass to the attention block for further processing. Thereafter a dropout is then applied and then for classification to respective parasite input image. The integration of dilated with attention mechanism is represented mathematically:

The input image is denoted as $x \in R_{HxWxC}$ where H W, and C representing the height, width, and number of channels of the image, respectively.

In the given architecture, there are three dilated convolutional layers, with each layer characterized by:

$$D_i = x * (w + b) \tag{1}$$

Where in this context, $w\ and\ b$ refer to the filters and bias, respectively.

Next, we employ dropout for regularization to mitigate overfitting. Following that, we introduce ReLU (Rectified Linear Unit) for non-linear purposes, allowing the model to capture complex relationships and patterns within the data.

$$D_{i(dropout)} = Dropout(D_i) \tag{2}$$

$$D_{i(ReLU)} = ReLU\left(D_{i(dropout)}\right) \tag{3}$$

Hence, it is essential to decrease the spatial dimension while preserving crucial features:

$$D_{i(max)} = Maxpooling\left(D_{i(ReLU)}\right) \tag{4}$$

Equations 1 to 4 are iteratively applied for each Dilated layer ($D_{i..n}$), where *i* is numbered 1 to 3 while *n* is the last convolution layer. Notably, each preceding layer extracts features that serve as inputs for the subsequent layer. In this sequence, Layer 1 serves as features for Layer 2, and Layer 2 serves as features for Layer 3.

Following the Dilated layers, the last one (D_3) is flattened, which means converting the three-dimensional tensor into a one-dimensional tensor. This transformation streamlines the data for further processing and preparing it for input into the dense layer.

$$D_{3(Flatten)} = Flatten\left(D_{i(max)}\right) \tag{5}$$

$$Flatten_{out} = D_{3(Flatten)} * (w + b) \tag{6}$$

After the dilated block output as shown in equation (6), Softmax operation is employed to generate attention weights, enabling the model to assign importance to different image features. The subsequent introduction of element-wise multiplication for self-attention enhances the model's focus on important features of the image, improving accuracy in malaria parasite classification by capturing complex shapes and relationships within the input data.

Thus, the flattened output is integrated into a self-attention mechanism block with a softmax:

$$Attn_w = Softmax(Flatten_{out}) \tag{7}$$

The output $Attn\_w$ is passed through a dense layer Dense layer with activation function ReLU, which gives:

$$Dense_{F\_out} = ReLU(Attn\_w) \tag{8}$$

Dropout is then applied for regularization purposes, serving to prevent overfitting by randomly deactivating a fraction of neurons during training. This helps enhance the model's generalization capabilities and robustness to unseen data. In the subsequent dense layer, features are processed using weights and biases. The introduction of weights allows the model to learn the importance of different features, while biases provide flexibility in adjusting the output. This weighted combination assists the model in capturing and representing complex relationships within the data.

$$Dropout_{out} = Dropout\left(Dense_{F_{out}}\right) \tag{9}$$

There after another dense layer was applied with ReLU as activation function:

$$Dense_{Layer\_out} = ReLU(Dropout_{out}) \tag{10}$$

Finally, the final dense layer ($Dense_{Layer\_out}$) processes features using weights and biases, and subsequently, the Sigmoid activation function is applied, in order for the output layer to transform the model's raw predictions into probability scores, facilitating binary classification into "infected" and "uninfected" classes. Thus, it facilitates the model's final decision-making process in a binary classification scenario.

$$Output\ _{Layer\_out} = Dense_{Layer\_out} * (w + b) \tag{11}$$

$$Final_{Output} = Sigmoid\left(Output\ _{Layer_{out}}\right) \tag{12}$$

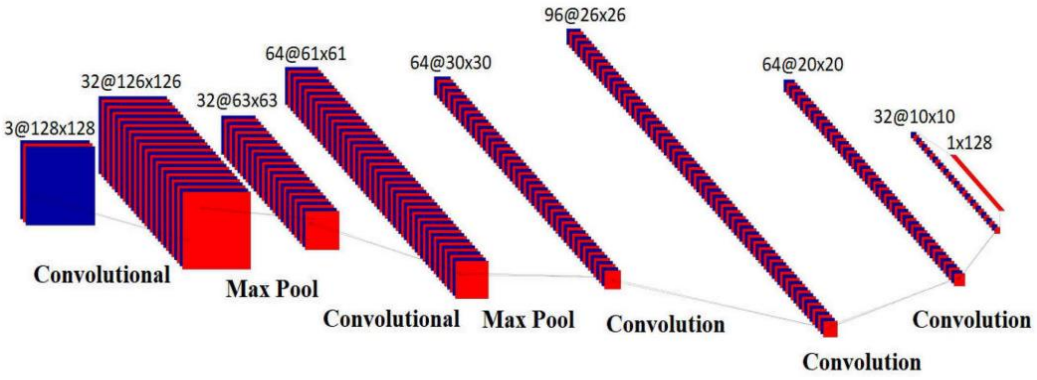


Fig. 4. CNN Architecture by [27]

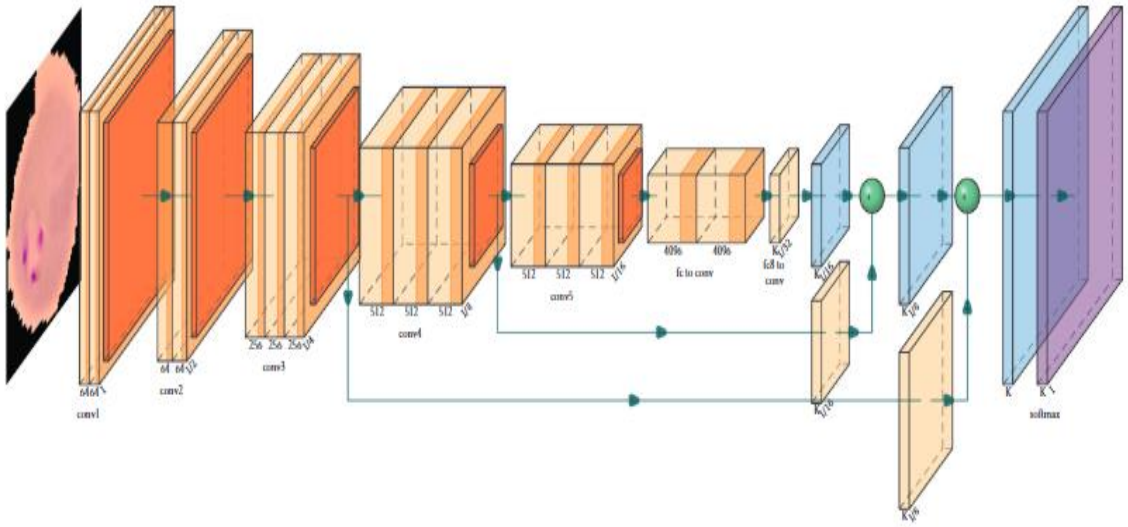


Fig. 5. CNN Model Architecture by [28]

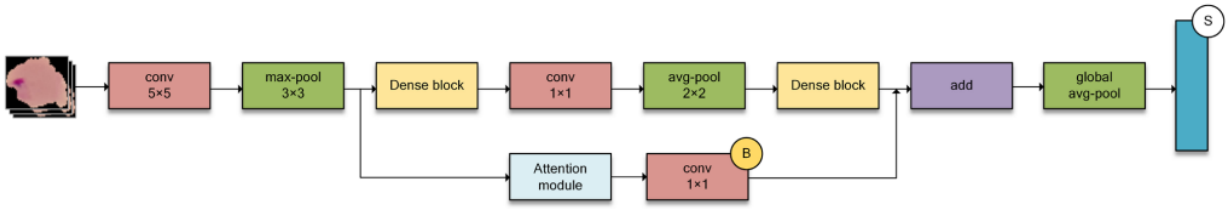


Fig. 6. Attentive Dense Circular Net (ADCN) based on CNN Architecture by [29]

4. Experimental Result Analysis

The proposed model underwent training utilizing malaria parasite datasets following image preprocessing, with 20% of the dataset reserved for testing the model's performance. The training process involved 10 epochs with a fixed learning rate of 0.0001, and batch size optimization was performed iteratively to enhance overall performance.

Various evaluation metrics were employed to gauge the deep learning model's effectiveness, including the Confusion Matrix (CM), accuracy, precision, recall, F1-score, and the Area Under the Curve (AUC).

Precision, denoting the likelihood that a positive call is genuinely an erythrocyte, is calculated using the formula:

$$Precision = \frac{TP}{TN+FP} \tag{13}$$

The F1 score, representing the harmonic mean of precision and sensitivity, serves as a measure of test accuracy in precision-recall evaluation:

$$F_1 = 2 \times \frac{Sensitivity \times Precision}{Sensitivity + Precision} \tag{14}$$

Accuracy, indicating the overall rate of correct predictions, is expressed mathematically as:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+TF} \tag{15}$$

AUC quantifies the overall ability of a model to discriminate between the positive and negative classes. Is a metric used to evaluate the performance of a binary classification model.

$$AUC = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TP}{TN+FP}\right) \tag{16}$$

4.1 Model parameters utilized in the assessment

We fine-tuned our model's hyperparameters to determine the optimal configuration for malaria parasite classification. This process involved systematic experimentation with various batch sizes, specifically 32, 64, 69, 74, 79, and 84. The entire model utilized the rectified linear unit (ReLU) as its main activation function, while the last layer we employed Softmax. For the loss function, we selected Cross-Entropy Loss. The dataset was partitioned into training and validation sets using an 80:20 ratio. To enhance the model's robustness and mitigate overfitting, we implemented a dropout rate of 0.2 as a regularization technique. Table 2 outlines the overall parameters used in the architecture.

Table 2. Model Parameters settings

Name	Parameters
Image Input Size	150 x 150
Learning rate	0.0001
Loss function	Cross-entropy
Optimizer	Adam
Epoch	10
Maxpooling	2,2
Patience	3
Activation Function	ReLU, Softmax
Dropout	0.2
Batch Size	Varied in sizes (32, 64, 74, 79 and 84)

4.2 Results obtained using different fine-tuned Batch size

In the results analysis, we conducted a thorough examination of the proposed model's effectiveness, focusing on the fine-tuning of various hyperparameters, with particular emphasis on the critical aspect of adjusting the batch size to attain a well-generalized outcome. As highlighted in the study, the choice of batch size significantly influences the training dynamics, where larger sizes expedite training but excessively large ones may compromise generalization.

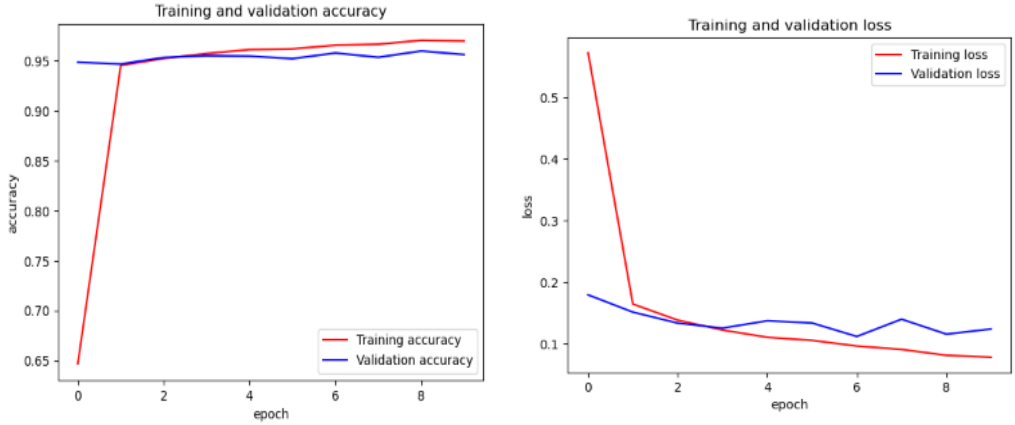
To address this, our study explored different batch sizes while keeping the epoch constant. Initially, batch sizes of 32 and 64 were examined, followed by subsequent increments of 5 until reaching a point of satisfactory performance. The variation in results across these different batch sizes provides valuable insights into the model's sensitivity to this hyperparameter.

Fig. 7 to fig. 12 present the loss and validation curves corresponding to each fine-tuned batch size, offering a comprehensive overview of the model's performance across varied batch sizes. These curves allow us to observe trends in convergence, generalization, and the trade-off between training dynamics and model robustness.

Additionally, to delve deeper into the model's predictive capabilities under different batch size configurations, we utilized confusion matrices as illustrated in Fig. 13 to fig. 18. These matrices provide detailed insights into the number of predicted values, enabling a better understanding of how the model performs in terms of correctly classifying infected and uninfected cells under various batch size scenarios.

The implications drawn from the examination of diverse batch sizes significantly shape our understanding of the results. They underscore the pivotal role of batch size selection in achieving high model performance, stressing the importance of adopting a subtle approach that balances effective training dynamics and robust generalization especially in the area of medical applications.

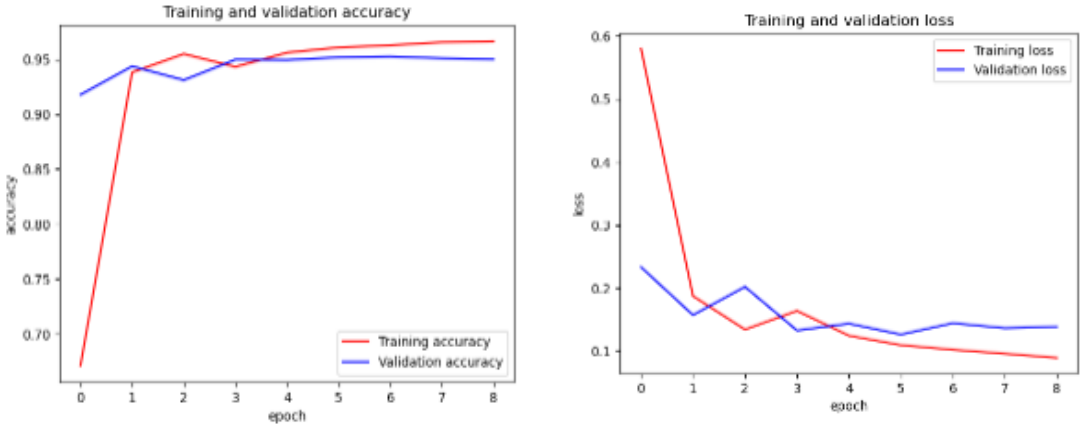
In Fig. 7(a & b), the model exhibits commendable performance, with training accuracy improving from 64.7% to approximately 97.0% by the tenth epoch, showcasing effective learning. Validation accuracy remains stable, starting at 94.9% and ending around 96.1%, emphasizing the model's generalization ability. Training loss decreases from 0.5721 to 0.0778, and validation loss decreases from 0.1790 to 0.1114, indicating refined fit and improved generalization. Overall, the model consistently improves in accuracy with decreasing losses, underscoring its effective learning and generalization.



(a) (b)

Fig. 7. (a) Training/validation accuracy for 32 batch size, (b) Training/validation loss for 32 batch size

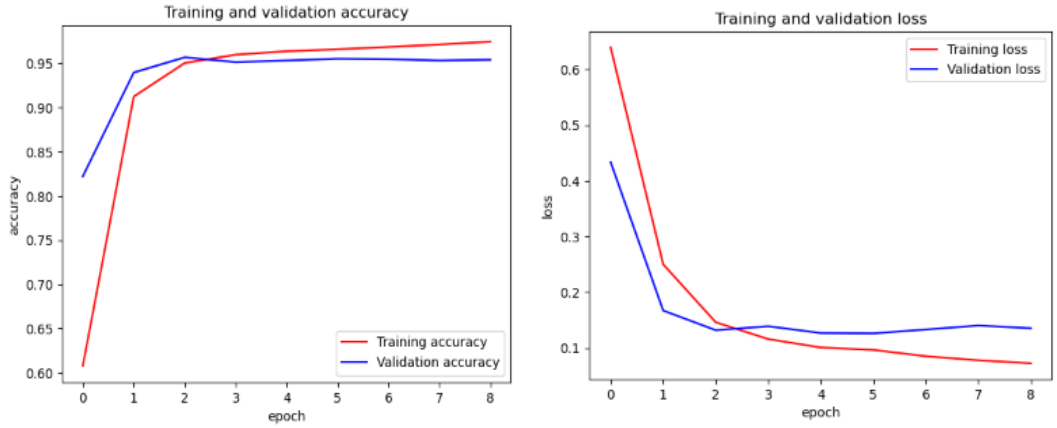
In Fig. 8(a & b), training accuracy starts at 67.1% and reaches around 96.7% by epoch 9, triggering early stoppage for overfitting prevention. Validation accuracy, beginning at 91.8%, fluctuates between epochs 2 to 9 but ends at approximately 95.0%, showcasing effective learning. Training loss drops from 0.5801 to 0.0778, reflecting progressive error minimization, while validation loss fluctuates but remains relatively low, reaching a minimum of 0.1114, indicating improved generalization. Caution is advised for potential overfitting in future epochs, warranting continuous monitoring.



(a) (b)

Fig. 8. (a) Training/validation accuracy for 64 batch size, (b) Training/validation loss for 64 batch size

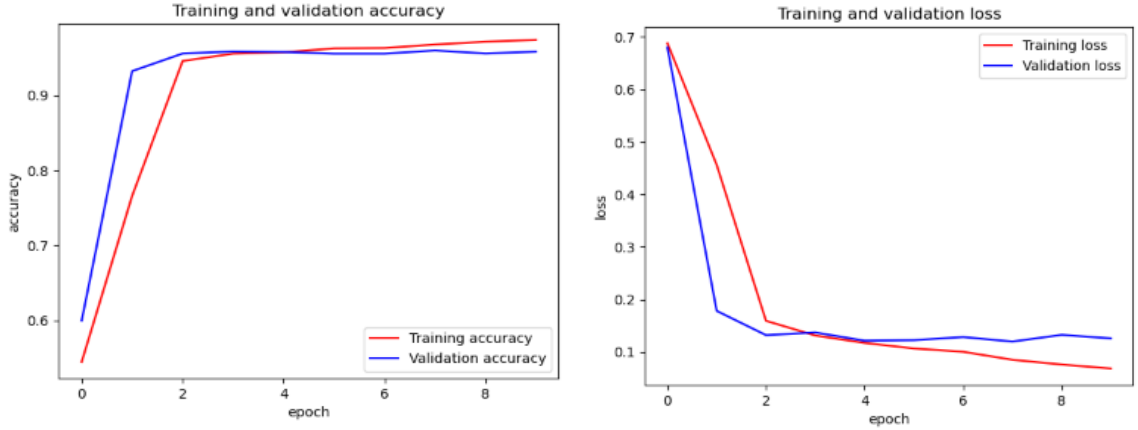
In Epoch 1 of Fig. 9(a & b), training accuracy starts at around 60.8%, steadily improving to about 97.4% from Epochs 2-10, showcasing effective learning. Validation accuracy, beginning at approximately 82.2% in Epoch 1, consistently improves to about 95.4% by Epochs 2-10, highlighting strong generalization. Training loss, initially at 0.6395 in Epoch 1, progressively decreases to a minimum of 0.0725 by the end of training, demonstrating progressive error minimization. Validation loss, starting at 0.4336 in Epoch 1, consistently decreases to a minimum of 0.1262 by Epochs 2-10, indicating improved generalization. Overall, both training and validation accuracy show consistent improvement, with decreasing losses emphasizing effective learning and fitting to the data.



(a) (b)

Fig. 9. (a) Training/validation accuracy for 69 batch size, (b) Training/validation loss for 69 batch size

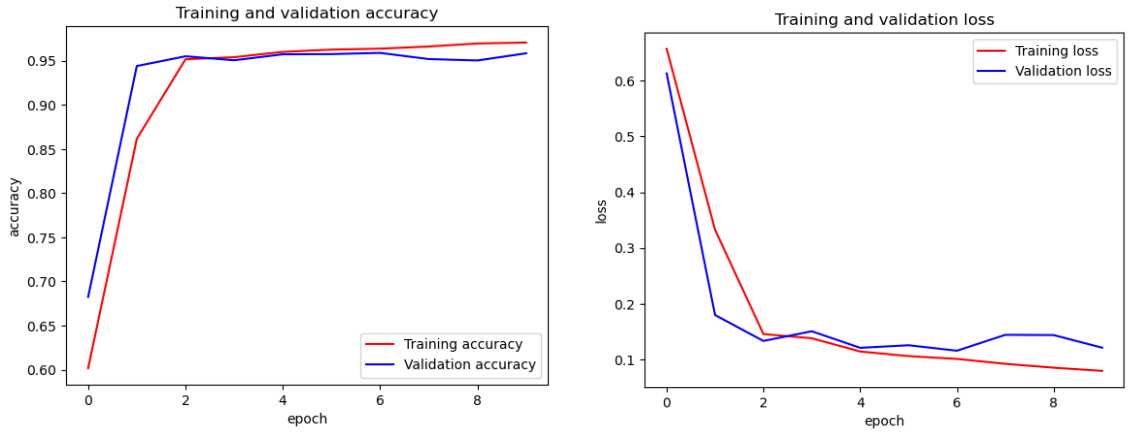
For batch size 69 in Fig. 10(a & b), training accuracy starts at 54.5% in Epoch 1, steadily improving to around 97.4% by Epochs 2-10, indicating effective learning. Validation accuracy, beginning at 59.9% in Epoch 1, continuously enhances between Epochs 2 to 10, concluding around 95.8%, highlighting effective generalization. The training loss, initially at 0.6881 in Epoch 1, steadily decreases to 0.0683 by the end of Epochs 2-10, signifying progressive error minimization. Simultaneously, validation loss, starting at 0.6800 in Epoch 1, consistently decreases during Epochs 2-10, reaching a minimum of 0.1196 by the end of training, indicating improved generalization. In summary, the model exhibits effective learning, with both training and validation accuracy consistently improving, and decreasing losses reflect its ability to fit the training data and generalize to unseen data.



(a) (b)

Fig. 10. (a) Training/validation accuracy for 74 batch size, (b) Training/validation loss for 74 batch size

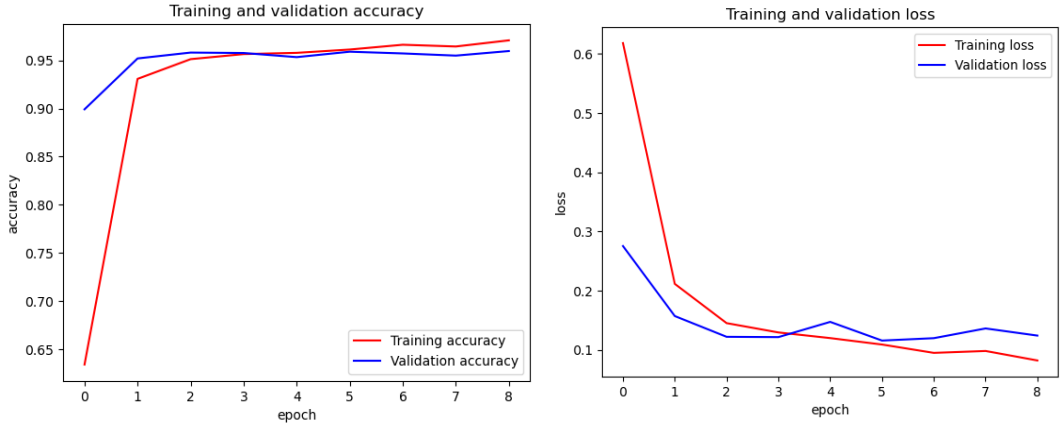
In Fig. 11(a & b), the model's learning trajectory is notable, with training accuracy beginning at 59.6% in Epoch 1 and steadily improving to approximately 96.9% by Epochs 2-10, indicating effective adaptation. Validation accuracy, starting at 78.5% in Epoch 1, consistently improves to around 95.8% by Epochs 2-10, showcasing effective generalization. The training loss, starting at 0.6632 in Epoch 1, decreases to 0.0876 by the end of Epochs 2-10, illustrating progressive error minimization. Similarly, validation loss, starting at 0.4865 in Epoch 1, consistently decreases to 0.1212 by the end of Epochs 2-10, indicating enhanced generalization. Overall, both training and validation accuracy show consistent improvement, alongside decreasing losses, emphasizing the model's effective learning, adaptation, and generalization.



(a) (b)

Fig. 11. (a) Training/validation accuracy for 79 batch size, (b) Training/validation loss for 79 batch size

In Epoch 1 of Fig. 12(a & b) for batch size 84, training accuracy starts at 55, substantially improving to around 97 between Epochs 2-9, signifying effective learning. Conversely, validation accuracy begins at 66.8, steadily improving to around 95.8 by the end of training, demonstrating generalization capacity. The training loss, starting at 0.5846 in Epoch 1, substantially decreases to a minimum of 0.0826 between Epochs 2-9, showcasing progressive error minimization. Validation loss, initially higher at 0.6442 in Epoch 1, consistently decreases to a minimum of 0.1215 by the end of training, indicating improved generalization. Both training and validation accuracy show consistent improvement, with the narrowing gap indicating avoidance of overfitting.



(a) (b)

Fig. 12. (a) Training/validation accuracy for 84 batch size, (b) Training/validation loss for 84 batch size

Furthermore, confusion matrices for all varied batch sizes were generated in order to quantify the performance metrics. The confusion matrices were generated using the testing dataset of 5540 images. The confusion matrices of each batch size is given in Fig. 13 to 18 which is summarise in table 3

Table 3. Summary of Confusion matrix results based on batch size

Batch Size	TP	FP	FN	TN
32	131	2555	2748	156
64	138	2582	2741	79
69	173	2585	2706	76
74	129	2563	2760	98
79	143	2554	2736	107
84	164	2595	2715	68

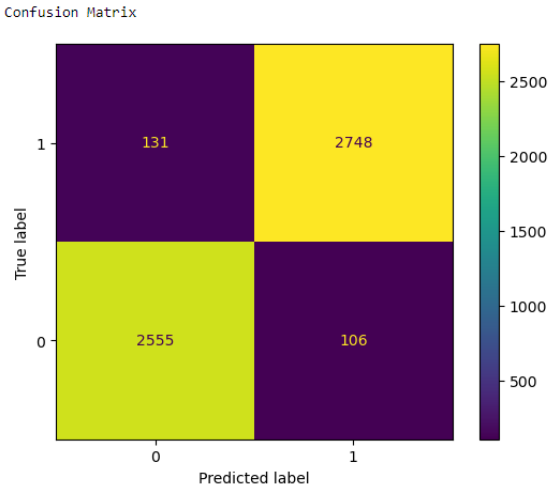


Fig. 13. Confusion Matrix for 32 batch size

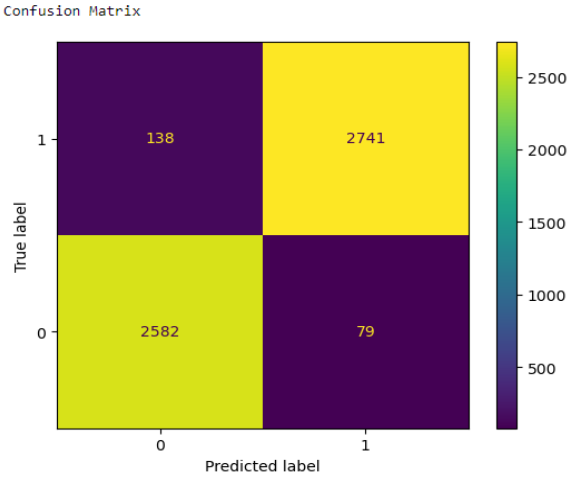


Fig. 14. Confusion Matrix for 64 batch size

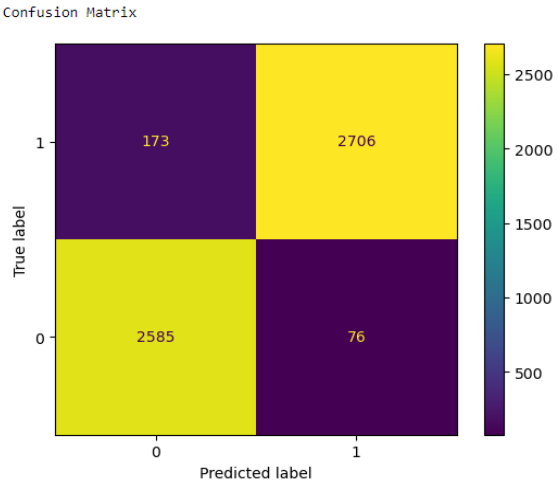


Fig. 15. Confusion Matrix for 69 batch size

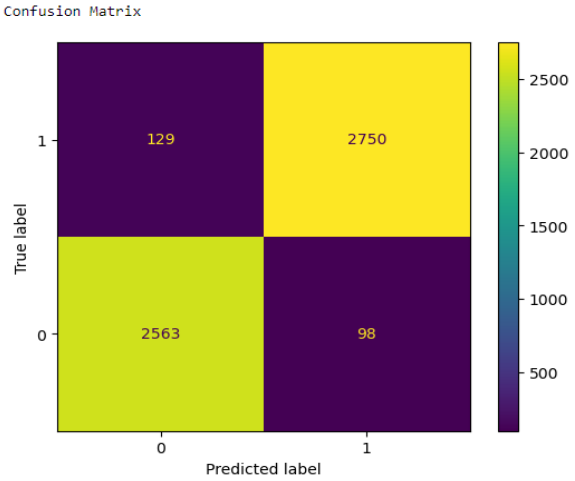


Fig. 16. Confusion Matrix for 74 batch size

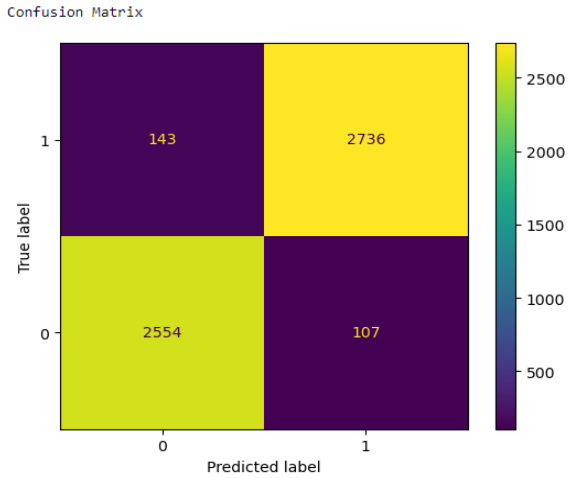


Fig. 17. Confusion Matrix for 79 batch size

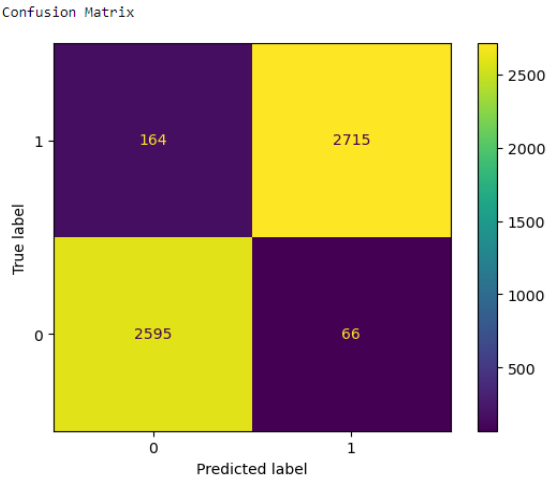


Fig. 18. Confusion Matrix for 84 batch size

Table 4 showcases the model performance across various batch sizes. Batch size 64 achieved the highest accuracy of 96.08 and AUC of 99.12. However, batch size 74 attained the highest recall value of 95.52, albeit with slightly lower accuracy and AUC compared to batch size 64. Meanwhile, batch size 32 yielded an accuracy of 95.72, with an AUC of 99.05.

Table 4. Model Performance based on incremental Batch size

Batch Size	Accuracy	Recall	F1_score	Precision	AUC
32	95.72	95.45	96	96.29	99.05
64	96.08	95.21	96	97.19	99.12
69	95.51	93.99	96	97.27	98.75
74	95.90	95.52	96	96.56	98.82
79	95.49	95.03	95	96.24	99.02
84	95.85	94.30	96	97.63	99.11

4.3 Comparing Performance with Previous Works

Utilizing the same datasets, our proposed model undergoes a comparative analysis with previous approaches, as outlined in Table 5. Notably, a majority of the prior works rely on CNN, which inherently possesses limitations in capturing the global context of image datasets. Consequently, these approaches are confined to local features, impacting their overall performance. In contrast, our work addresses this limitation by encompassing a broader range of image features and employing a larger batch size, striking a balance between speed and computational efficiency. This choice also contributes to a more stable parameter update for the model.

In our experiments, the model achieved a precision of 97.63% for batch size 64, outperforming previous work [26], though [29] reported a higher precision of 98.79%. However, we found a trade-off with recall, as our model had a recall of 93.99%, which was lower than the 96.33% reported in [26]. This implies that, while the model reduces false positives (high precision), it may overlook some true positives (lower recall). In terms of accuracy, [29] scored the highest at 98.85%, followed by [30], and our model scored 96.08%. Furthermore, the model's AUC score of 99.12 with a batch size of 64 demonstrates its ability to distinguish between infected and uninfected cases.

This study sheds light on the limitations of conventional CNN models, rooted in local receptive fields that generalize or classify based on image patterns. In response, our work incorporates Dilated CNN, essential for extending CNN models with a larger kernel to capture the global information of malaria images. The integration of an attention mechanism with Dilated CNN enhances our model's capability to focus on fine details or specific features, providing flexibility in attending to various parts of the input with varying weights. While Dilated CNN may struggle with coarse and fine-grained details uniformly, the attention mechanism overcomes these limitations. By synergizing these models, our proposed approach combines the strengths of each to create a potent model capable of capturing both local and global characteristics of the malaria parasite, ensuring exceptional coverage and performance.

Table 5. Comparison of previous work with the proposed model

Author	Model Used	Precision	Recall	F1-Score	Accuracy	AUC
[26]	Dilated CNN	95.80	96.33	96.06	96.05	Nil
[27]	ADCN Based CNN	Nil	Nil	97.50	97.47	Nil
[29]	CNN	98.79	NIL	Nil	98.85	Nil
[30]	CNN+ Spatial+ Channel attention module	Nil	Nil	Nil	97.80	Nil
[Proposed Model]	Dilated CNN +Attention Mechanism	97.63	93.99	96	96.08	99.12

The model's ability to decrease false positives is demonstrated by the reported precision score of 97.63%; however, concentrating only on precision ignores the significance of recall and F1-score, which are essential for evaluating the model's overall performance and capacity to identify true positives. In order to address this, we will provide a balanced assessment of the model's efficacy in identifying all positive cases by including thorough recall and F1-score evaluations for each batch size.

5. Conclusion

The observed variations in results across different batch sizes in our study underscore the model's sensitivity to the choice of this hyperparameter in the classification of malaria parasites. This highlights the critical importance of achieving a balance to optimize performance on unseen data, especially in the context of medical applications where the reliability of the model is paramount for accurate diagnosis and treatment decisions. The model's classification performance also proves pivotal in supporting clinical decision-making processes.

In conclusion, our research has significantly contributed to advancing malaria diagnosis by integrating attention mechanisms and dilated CNN architectures. Through rigorous testing across multiple batch sizes, we identified the model's optimal performance with an accuracy of 96.08, recall of 93.99, F1-score of 96, precision of 97.63, and a noteworthy AUC of 99.12, achieved with a batch size of 64. These outcomes underscore the critical importance of hyperparameter choices, emphasizing the need for a balanced approach to ensure the model's reliability. This sensitivity

has far-reaching implications for the medical field, where accurate and dependable diagnostic tools are essential for effective clinical decision-making.

As part of future work, we aim to further enhance the robustness and applicability of our model by exploring its potential for categorizing different species of malaria parasites. Leveraging datasets specific to various species will allow us to extend the model's capabilities and address diverse challenges in malaria diagnosis. This ongoing exploration will contribute to the continued evolution of diagnostic tool, in ensuring its relevance and effectiveness in diverse clinical scenarios. In addition, reliance on a single NIH dataset raises concerns about the model's generalizability, so future work should also validate it with diverse malaria datasets and employ transfer learning to enhance robustness and reliability in real-world detection scenarios.

References

[1] WHO (World Health Organization). Malaria microscopy quality assurance manual – Version. 2. WHO, page 140, 2016.

[2] Acevedo, A., Alférez, S., Merino, A., Puigví, L., & Rodellar, J. (2019). Recognition of peripheral blood cell images using convolutional neural networks. *Computer methods and programs in biomedicine*, *180*, 105020.

[3] Rosado, L., Da Costa, J. M. C., Elias, D., & Cardoso, J. S. (2017). Mobile-based analysis of malaria-infected thin blood smears: automated species and life cycle stage determination. *Sensors*, *17*(10), 2167.

[4] Bain, B. J. (2005). Diagnosis from the blood smear. *New England Journal of Medicine*, *353*(5), 498-507.

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[6] Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2020). Local features and global shape information in object classification by deep convolutional neural networks. *Vision research*, *172*, 46-61.

[7] Liu, Y. H. (2018, September). Feature extraction and image recognition with convolutional neural networks. In *Journal of Physics: Conference Series* (Vol. 1087, p. 062032). IOP Publishing.

[8] Purnama, I. K. E., Rahmanti, F. Z., & Purnomo, M. H. (2013, November). Malaria parasite identification on thick blood film using genetic programming. In *2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME)* (pp. 194-198). IEEE.

[9] Bibin, D., Nair, M. S., & Punitha, P. (2017). Malaria parasite detection from peripheral blood smear images using deep belief networks. *IEEE Access*, *5*, 9099-9108.

[10] Rajaraman, S., Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., ... & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, *6*, e4568.

[11] Sriporn, K., Tsai, C. F., Tsai, C. E., & Wang, P. (2020). Analyzing malaria disease using effective deep learning approach. *Diagnostics*, *10*(10), 744.

[12] Goni, M. O. F., Mondal, M. N. I., Islam, S. R., Nahiduzzaman, M., Islam, M. R., Anower, M. S., & Kwak, K. S. (2023). Diagnosis of Malaria Using Double Hidden Layer Extreme Learning Machine Algorithm With CNN Feature Extraction and Parasite Inflator. *IEEE Access*, *11*, 4117-4130.

[13] Khan, A., Gupta, K. D., Venugopal, D., & Kumar, N. (2020, July). Cidmp: Completely interpretable detection of malaria parasite in red blood cells using lower-dimensional feature space. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

[14] Fuhad, K. F., Tuba, J. F., Sarker, M. R. A., Momen, S., Mohammed, N., & Rahman, T. (2020). Deep learning-based automatic malaria parasite detection from blood smear and its smartphone-based application. *Diagnostics*, *10*(5), 329.

[15] Islam, M. R., Nahiduzzaman, M., Goni, M. O. F., Sayeed, A., Anower, M. S., Ahsan, M., & Haider, J. (2022). Explainable transformer-based deep learning model for the detection of malaria parasites from blood cell images. *Sensors*, *22*(12), 4358.

[16] Mohanty, I., Pattanaik, P. A., & Swarnkar, T. (2019). Automatic detection of malaria parasites using unsupervised techniques. In *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)* (pp. 41-49). Springer International Publishing.

[17] Dong, Y., Jiang, Z., Shen, H., Pan, W. D., Williams, L. A., Reddy, V. V., ... & Bryan, A. W. (2017, February). Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In *2017 IEEE EMBS international conference on biomedical & health informatics (BHI)* (pp. 101-104). IEEE..

[18] Roy, K., Sharmin, S., Mukta, R. M., & Sen, A. (2018). Detection of malaria parasite in Giemsa blood sample using image processing. *International Journal of Computer Science and Information Technology*, *10*(1), 55-65.

[19] Li, D., & Ma, Z. (2022). Residual attention learning network and SVM for malaria parasite detection. *Multimedia Tools and Applications*, *81*(8), 10935-10960.

[20] Manning, K., Zhai, X., & Yu, W. (2022). Image analysis and machine learning-based malaria assessment system. *Digital Communications and Networks*, *8*(2), 132-142.

[21] Reddy, A. S. B., & Juliet, D. S. (2019, April). Transfer learning with ResNet-50 for malaria cell-image classification. In *2019 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0945-0949). IEEE.

[22] Krishnadas, P., & Sampathila, N. (2021, July). Automated detection of malaria implemented by deep learning in PyTorch. In *2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (pp. 01-05). IEEE.

[23] Taha, B., & Liza, F. R. (2021, December). Automatic identification of malaria-infected cells using deep convolutional neural network. In *2021 24th International Conference on Computer and Information Technology (ICCIT)* (pp. 1-5). IEEE..

[24] Maqsood, A., Farid, M. S., Khan, M. H., & Grzegorzek, M. (2021). Deep malaria parasite detection in thin blood smear microscopic images. *Applied Sciences*, *11*(5), 2284.

[25] Shal, A., & Gupta, R. (2022, January). A comparative study on malaria cell detection using computer vision. In *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 548-552). IEEE.

[26] Gummadi, S. D., Ghosh, A., & Vootla, Y. (2022, June). Transfer Learning based Classification of Plasmodium Falciparum Parasitic Blood Smear Images. In *2022 10th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-5). IEEE.
[27] Qayyum, A. B. A., Islam, T., & Haque, M. A. (2019, November). Malaria diagnosis with dilated convolutional neural network based image analysis. In *2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON)* (pp. 68-72). IEEE.
[28] Alharbi, A. H., Aravinda, C. V., Shetty, J., Jabarulla, M. Y., Sudeepa, K. B., & Singh, S. K. (2022). Computational models-based detection of peripheral malarial parasites in blood smears. *Contrast Media & Molecular Imaging*, *2022*.
[29] Quan, Q., Wang, J., & Liu, L. (2020). An effective convolutional neural network for classifying red blood cells in malaria diseases. *Interdisciplinary Sciences: Computational Life Sciences*, *12*, 217-225.
[30] Fu, M., Wu, K., Li, Y., Luo, L., Huang, W., & Zhang, Q. (2023). An intelligent detection method for plasmodium based on self-supervised learning and attention mechanism. *Frontiers in Medicine*, *10*.

Authors' Profiles

Suleiman Garba is a staff of FCT College of Education Zuba, Abuja, Computer Science Department. He holds Masters (M.Tech) in Computer Science from Federal University of Technology, Minna in 2015 and currently a PhD student of Computer Science Federal University of Technology, Minna. He had been teaching for over 7 years in Secondary School before joining FCT College of Education, Zuba, Abuja-Nigeria in 2007. He is currently a Ph.D. student of Computer Science Department, Federal University of Technology, Minna, Niger State.

Muhammad Bashir Abdullahi received his B.Tech (Honors) in Mathematics/Computer Science from Federal University of Technology, Minna-Nigeria and Ph.D. in Computer Science and Technology from Central South University, Changsha, Hunan, P. R. China. He was former Head of Department, Computer Science, Federal University of Technology, Minna. His current research interests include trust, security and privacy issues in data management for wireless sensor and ad hoc networks, Cloud computing, Big data technology and information and communication security. He is a member of Computer Professionals (Registration Council) of Nigeria (CPN) and Nigeria Computer Society (NCS).

Sulaimon Adebayo Bashir is a Senior Lecturer in the Department of Computer Science, Federal University of Technology, Minna, Nigeria. He received PhD degree in Computing from Robert Gordon University Aberdeen UK in 2017, MSc. in Computer Science from the Nigeria premier university, University of Ibadan Nigeria in 2008 and a Bachelor of Technology in Computer Science from Ladoke Akintola University of Technology Ogbomoso, Nigeria in 2003. He is a recipient of the National Information Technology Development Agency Scholarship (2012). He is a member of the ACM, the Computer Professional Registration Council of Nigeria and Nigeria Computer Society. He has headed Computer Science Department, Federal University of Technology, Minna. His research interests include Machine Learning and its application to Activity Recognition, Computer Vision, Cyber Security, Social Media Mining and Intelligent Healthcare Systems. He teaches and supervises both undergraduate and postgraduate MSc. and PhD students and has numerous publications to his credit.

Abisoye Opeyemi Aderike is currently a Senior Lecturer and the Head of the Department of Computer Science, Federal University of Technology, Minna, Niger State, Nigeria from 2021 till Date. She attended University of Ilorin, Nigeria, where she got Ph.D. in Computer Science. She has over Fifty (50) publications in both reputable international, local journals and conference proceedings. She is a member of Computer Professional Registration Council of Nigeria (MPCN) and a member of Nigeria Computer Society (NCS). She has served as Departmental Postgraduate Coordinator from 2017 till 2021. She has also served as Turnitin Officer for the School of Information and Communication Technology from 2014 till 2017 and Head of the Computer Science Department, Federal University of Technology, Minna. She is supervising/co-supervising over Twenty (20) students in both Masters and PhD Level.

How to cite this paper: Suleiman Garba, Muhammad Bashir Abdullahi, Sulaimon Adebayo Bashir, Abisoye Opeyemi Aderike, "Dilated Convolutional Neural Network with Attention Mechanism for Classification of Malaria Parasites", International Journal of Engineering and Manufacturing (IJEM), Vol.14, No.6, pp. 11-26, 2024. DOI:10.5815/ijem.2024.06.02