



Deep Learning Approaches for Malaria Diagnosis: A Comparative Study of Custom CNN and Transfer Learning Models in Blood Smear Analysis

Sıtma Tanısı İçin Derin Öğrenme Yaklaşımları: Kan
Sürüntüsü Analizinde Özel CNN ve Transfer Öğrenme
Modellerinin Karşılaştırmalı Çalışması

Anselme ATCHOGOU¹, Abdoul MALİK²

¹Department of Intelligent Systems Engineering, Engineering Faculty, Ondokuz Mayıs University, Samsun, Türkiye
· anselmeatchogou@gmail.com · ORCID > 0009-0002-1593-516X

²Department of Intelligent Systems Engineering, Engineering Faculty, Ondokuz Mayıs University, Samsun, Türkiye
· aabdoulmalik99@gmail.com · ORCID > 0009-0006-6525-8346

Makale Bilgisi/Article Information

Makale Türü/Article Types: Araştırma Makalesi/Research Article

Geliş Tarihi/Received: 9 Ocak/January 2025

Kabul Tarihi/Accepted: 13 Ekim/October 2025

Yıl/Year: 2025 | **Cilt-Volume:** 5 | **Sayı-Issue:** 2 | **Sayfa/Pages:** 51-71

Atıf/Cite as: Atchogou, A., Malik, A. "Deep Learning Approaches for Malaria Diagnosis: A Comparative Study of Custom CNN and Transfer Learning Models in Blood Smear Analysis" Ondokuz Mayıs University Journal of Engineering Sciences and Technology 5(2), Eylül 2025: 51-71.

Sorumlu Yazar/Corresponding Author: Abdoul MALİK

DEEP LEARNING APPROACHES FOR MALARIA DIAGNOSIS: A COMPARATIVE STUDY OF CUSTOM CNN AND TRANSFER LEARNING MODELS IN BLOOD SMEAR ANALYSIS

ABSTRACT

The Plasmodium genus of single-celled parasites is the cause of malaria. These parasites are spread to humans through the bite of an infected Anopheles mosquito. Particularly in sub-Saharan Africa, where it severely strains health systems and economics, the illness remains an essential public health issue. Effective diagnosis and treatment are crucial for controlling and eventually eliminating malaria. The microscopic analysis of blood smears is the conventional yet labour-intensive method for diagnosing malaria, demanding extensive expertise. Automated detection through deep learning presents a vital alternative, particularly crucial for sub-Saharan Africa. This study aims to meticulously compare the performance of a custom-designed Convolutional Neural Network (CNN) with five advanced transfer learning models, ResNet50, VGG19, InceptionV3, EfficientNet-B3, EfficientNet-B7, and YOLOv11m, in categorizing segmented red blood cell images for malaria detection. Our approach involves comprehensive image preprocessing, data augmentation, and the implementation of various models. The models were evaluated using a National Library of Medicine dataset based on various metrics, including F1 Score, Accuracy, Precision, Recall, Matthews Correlation Coefficient (MCC), AUC ROC, and AUC PR. The EfficientNet-B3 model emerged as the top performer, surpassing even the custom CNN with an impressive F1 Score of 98.12%, Accuracy of 98.08%, and an MCC of 96.15%, demonstrating its superior predictive power and reliability. YOLOv11m also showed strong performance with an F1 Score of 96.89%, Accuracy of 96.93%, and MCC of 93.91%, highlighting its efficiency for real-time applications. Although the custom CNN did not outperform the advanced models, it still exhibited commendable performance, underscoring the potential of tailored architectures. The results of this study demonstrate the great potential that deep learning methods have to improve the precision of malaria diagnosis, providing notable benefits to the healthcare systems, especially for regions most severely impacted by the disease.

Keywords: Malaria Diagnosis, Deep Learning, CNN, Transfer Learning, Healthcare.



SITMA TANISI İÇİN DERİN ÖĞRENME YAKLAŞIMLARI: KAN SÜRÜNTÜSÜ ANALİZİNDE ÖZEL CNN VE TRANSFER ÖĞRENME MODELLERİNİN KARŞILAŞTIRMALI ÇALIŞMASI

ÖZ

Tek hücreli parazitlerin Plasmodium cinsi, sıtmanın nedenidir. Bu parazitler, enfekte olmuş Anopheles sivrisineklerinin ısırmasıyla insanlara bulaşır. Özellikle sağlık sistemlerini ve ekonomiyi ciddi şekilde zorlayan Sahra altı Afrika'da, bu hastalık önemli bir halk sağlığı sorunu olmaya devam etmektedir. Sıtmayı kontrol altına almak ve nihayetinde ortadan kaldırmak için etkili tanı ve tedavi çok önemlidir. Kan yaymalarının mikroskopik analizi, sıtmayı teşhis etmek için geleneksel ancak emek yoğun bir yöntemdir ve kapsamlı uzmanlık gerektirir. Derin öğrenme yoluyla otomatik tespit, özellikle Sahra altı Afrika için hayati bir alternatif sunmaktadır. Bu çalışma, sıtma tespiti için segmentlere ayrılmış kırmızı kan hücresi görüntülerini sınıflandırmada, özel olarak tasarlanmış bir Convolutional Neural Network (CNN) ile beş gelişmiş transfer öğrenme modeli olan ResNet50, VGG19, InceptionV3, EfficientNet-B3, EfficientNet-B7 ve YOLOv11m'nin performansını titizlikle karşılaştırmayı amaçlamaktadır. Yaklaşımımız, kapsamlı görüntü ön işleme, veri artırma ve çeşitli modellerin uygulanmasını içermektedir. Modeller, F1 Skoru, Doğruluk, Hassasiyet, Geri Çağırma, Matthews Korelasyon Katsayısı (MCC), AUC ROC ve AUC PR gibi çeşitli metriklerle dayalı olarak Ulusal Tıp Kütüphanesi veri seti kullanılarak değerlendirildi. EfficientNet-B3 modeli, %98,12'lik etkileyici bir F1 Skoru, %98,08'lik Doğruluk ve %96,15'lik MCC ile özel CNN'yi bile geride bırakarak en iyi performans gösteren model olarak öne çıktı ve üstün tahmin gücü ve güvenilirliğini kanıtladı. YOLOv11m de %96,89 F1 Skoru, %96,93 Doğruluk ve %93,91 MCC ile güçlü bir performans sergiledi ve gerçek zamanlı uygulamalar için verimliliğini vurguladı. Özel CNN, gelişmiş modelleri geride bırakmasa da, yine de övgüye değer bir performans sergiledi ve özel olarak tasarlanmış mimarilerin potansiyelini vurguladı. Bu çalışmanın sonuçları, derin öğrenme yöntemlerinin sıtma tanısının doğruluğunu artırma konusunda büyük potansiyele sahip olduğunu ve özellikle bu hastalıktan en ciddi şekilde etkilenen bölgelerde sağlık sistemlerine önemli faydalar sağladığını göstermektedir.

Anahtar Kelimeler: Sıtma Tanısı, Derin Öğrenme, CNN, Transfer Öğrenme, Sağlık Hizmetleri.



Highlights

- EfficientNet-B3 achieved the highest performance with an F1 Score of 98.12% and an MCC of 96.15%.
- Custom CNN showed promise with tailored architectures for malaria detection.
- Six advanced transfer learning models, including ResNet50 , Efficient-Net-B7, and YOLOv11m, were compared.
- Comprehensive preprocessing and data augmentation enhanced diagnostic accuracy.
- Study highlights the potential of deep learning for improving malaria diagnosis.

1. INTRODUCTION

Malaria is an acute febrile illness predominantly afflicting tropical and subtropical regions. Its transmission cycle is facilitated by the *Anopheles* mosquito, which serves as a vector for the *Plasmodium* parasite that is injected into humans during a mosquito's blood meal [1]. The pathology of malaria involves a complex lifecycle with the parasites multiplying in the liver before invading red blood cells, often leading to severe clinical manifestations [2]. Five *Plasmodium* species infect humans, with *P. falciparum* and *P. vivax* accounting for most global malaria morbidity [3]. The World Health Organization reports a staggering 229 million cases worldwide in 2019, with a significant mortality burden [4]. In regions where malaria is endemic, such as sub-Saharan Africa, the disease's prevalence is compounded by insufficient healthcare infrastructure and a shortage of skilled diagnosticians [5]. The global health community, driven by initiatives like the United Nations and the Gates Foundation, aims to eradicate malaria by 2040 [6]. The realization of these goals hinges on the advancement of rapid, accurate diagnostic modalities.

While adequate, traditional microscopic examination of stained blood smears is time-consuming and relies heavily on pathologists' expertise [7]. Consequently, there is a critical need for innovative diagnostic solutions, especially in resource-limited settings. To address this issue, research has pivoted towards leveraging computer vision, machine learning, and deep learning techniques, which have shown promise in automating the detection of malarial parasites [8]. Convolutional Neural Networks (CNNs), Transfer Learning, and Model Ensembles represent the cutting edge of such technologies, demonstrating capabilities that rival human expertise.

This study aims to evaluate the effectiveness of a custom CNN model in comparison to preeminent transfer learning architectures ResNet50, VGG19, InceptionV3, EfficientNet-B3, EfficientNet-B7, and YOLOv11m in the classification of segmented blood cell images for malaria detection. The comparative analysis employs an array of performance metrics, including F1 Score, Accuracy, Precision, Recall, Matthews Correlation Coefficient (MCC), AUC ROC, and AUC PR, using a dataset provided by the National Library of Medicine. Our results illuminate the potential of deep learning applications to significantly improve malaria diagnostics, potentially offering life-saving support in the most affected regions.

The paper is structured into several main parts. It begins with a 'Literature Review' which places the study within the broader research field. Next, the 'Materials and Methods' section thoroughly explains the Dataset, System Design, and Model Architectures. The 'Results and Discussion' section then interprets the study's findings. Finally, the 'Conclusion' section succinctly wraps up the research and suggests avenues for further investigation.

2. RELATED WORKS

Machine learning (ML) and deep learning (DL) approaches have entirely changed the search for an efficient method of detecting malaria. Traditional ML methods hinged on manual feature extraction, drawing from domain knowledge of blood smear morphology and parasitic life cycles. Savkare and Narote [9], Bairagi and Charpe [10], and others have employed a variety of features ranging from textual and morphological to statistical, feeding these into algorithms like SVMs and AdaBoost for classification tasks with notable successes.

The inception of deep learning has shifted the paradigm from manual to automatic feature learning. CNNs, for example, have automated feature extraction processes, leading to significant breakthroughs in accuracy and efficiency. Maqsood et al. [11] achieved an accuracy of 96.82% with CNNs enhanced by image augmentation. Rajaraman et al. [8] attained high specificity and sensitivity using a pre-trained ResNet50 model, demonstrating the effectiveness of transfer learning in this field. Farah and Ammar (2023) proposed a basic CNN model consisting of 3 convolution blocks followed by fully connected layers. They compared the proposed CNN model with other pre-trained models, including VGG-19, ResNet50, DenseNet121, and Inception V3. The results of their proposed CNN achieved an accuracy score of 97% [12]. Irmak (2021) presented a pioneering deep-learning technique for malaria disease detection, employing a Convolutional Neural Network (CNN) with 20 weighted layers. The model achieved an impressive overall accuracy score of 95.28% [13].

Krishnadas and Sampathila (2021) focus on employing deep learning techniques implemented in PyTorch, mainly transfer learning, for malaria detection in segmented red blood cell images. Their method consists of using pre-trained ImageNet models (ResNet, DenseNet, etc.) and fine-tuning them to identify infected (parasitized) or uninfected cell pictures. Notably, they report that the DenseNet121 model achieved the highest accuracy of 94.43% in this task [14]. In their work published in 2023, Omar Faruq Goni et al. introduced an innovative approach for malaria prognosis utilizing the Extreme Learning Machine (ELM) algorithm. Their method incorporates Convolutional Neural Networks (CNN), ELM, and double hidden layer (DELM) as classifiers. Notably, their proposed CNN-DELM method achieved an impressive accuracy rate of 97.79% [15]. In their 2023 study, Al-muhaya et al. [16] compared the performance of four CNN models (GoogLeNet, DenseNet161, MobileNet_v2, and ResNet18) in detecting malaria using the publicly available Malaria Cell Images Dataset from NIH. Their results revealed that DenseNet161 achieved the highest accuracy, reaching 95.86%.

Agrawal et al. (2024) employed the Malaria Cell Images Dataset from Kaggle to assess the efficacy of semi-supervised learning methods in attaining high accuracy despite having limited labelled data. Remarkably, their approach achieved an impressive accuracy rate of 96% [17]. Nakasi et al. [18] explored approaches such as Faster-RCNN for object detection in malaria diagnosis, achieving an accuracy of 93.03%. Quan et al (2020) proposed the Attentive Dense Circular Net (ADCN), a classification model based on Convolutional Neural Networks (CNN), combining the principles of residual and dense networks with an attention mechanism. Trained and evaluated on a public dataset of red blood cell images, the model was compared with several established CNN architectures. The results show a clear superiority of ADCN, achieving an accuracy of 97.47%, a sensitivity of 97.86% and a specificity of 97.07%, systematically outperforming the best reference models [19].

Magotra and Rohil (2022) proposed two pre-trained CNN models using VGG-16 and Inception V3. Their results achieve 96% accuracy [20]. According to Pimple et al. (2022), CNN models are more efficient than Feed-Forward Neural Networks (FFNs). They advocate for using CNNs, citing that the filters utilized in CNNs extract image patterns while reducing data transfer across layers, thereby enhancing efficiency. The author proposed a CNN model and trained it using malaria images. Their CNN model achieved an impressive accuracy score of 94.52% [21]. In their study in 2016, Liang et al. (2016) introduced a machine-learning technique employing a CNN to categorize individual cells in thin films autonomously [22]. Using a unique 16-layer CNN model on cell pictures, scientists achieved an average accuracy of 97.37% during ten-fold cross-validation on a 27,578 single-cell dataset.

Recent studies have incorporated advanced object detection models like YOLO for malaria parasite detection. For instance, this paper [23] proposed a Fine-Tuned

YOLO-Based Deep Learning Model for Detecting Malaria Parasites and Leukocytes in Thick Smear Images. The findings from these studies highlight YOLO's potential for fast and accurate detection in diverse imaging conditions. In conclusion, the literature presents a trajectory of continuous improvement in malaria detection techniques. As computational power grows and machine learning algorithms become more refined, the capability of these models to accurately identify malaria from cell images continues to advance, offering promising implications for global health, especially in regions most affected by the disease.

3. MATERIALS AND METHODS

This study used Google Colab [24], which has an Nvidia Tesla T4 GPU and 32 GB of RAM for computational work. Python 3.10 was used for programming via the Jupyter interface. The OpenCV library was used for image processing, and Keras [25], with a TensorFlow [26] backend, was used to create deep-learning models. Regarding YOLOv11m, the Ultralytics YOLO library [27] was used.

The system architecture diagram below (Figure 1) illustrates the workflow adopted in our study, which encompasses data procurement to the final classification outcomes. This schematic visualizes the methodical procedure to evaluate the efficacy of various deep-learning models in identifying malaria from blood smear images.

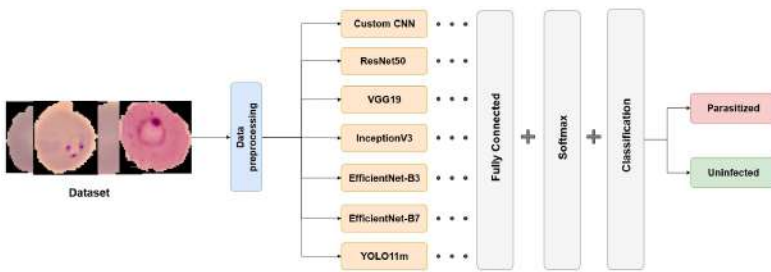


Figure 1. System Architecture Diagram

3.1. Dataset

This study utilizes the “National Library of Medicine” dataset, available on Kaggle [28], which is part of the “National Institutes of Health” collection in the United States. The dataset comprises segmented red blood cell images derived from Giemsa-stained slides, contributed by 150 infected and 50 uninfected individuals [8]. This balanced dataset contains 27,558 erythrocyte images, evenly split between parasitized and non-parasitized groups, as demonstrated in Figure 2 with examples from both categories.

The dataset is divided into three sets: training, validation, and testing. The training set includes 80% of the data. A subset of the training data, 20%, is reserved for validation to fine-tune the model parameters. The remaining 20% of the original dataset is set aside as the test set, which is used to evaluate the model's performance on data it has never seen before. These proportions ensure that the model has a substantial amount of data for learning, a sufficient number of images for validation during the learning process, and a separate set for final performance evaluation. Figure 3 shows samples distributed according to the categories.

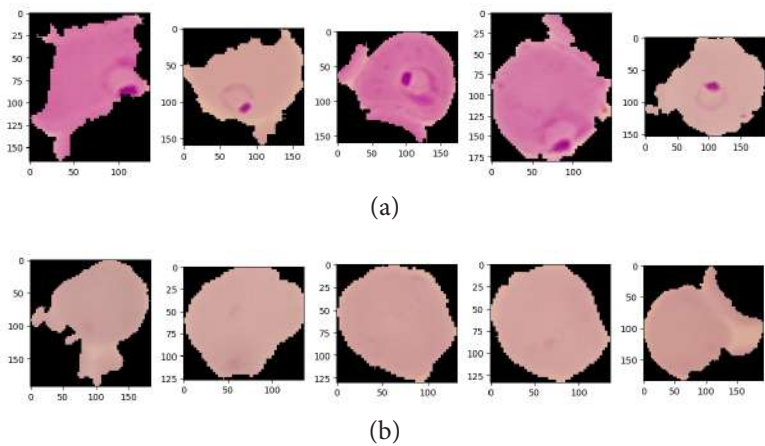


Figure 2. (a) Displays Parasitized Images, and (b) Uninfected Images

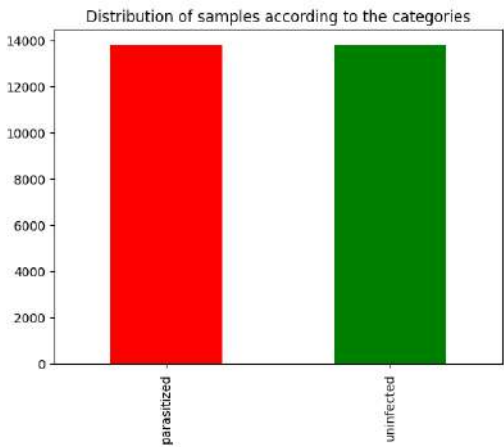


Figure 3. Samples distributed according to the classes

3.2. Data Preprocessing

- **Image Color (RGB) Distribution and Histograms**

Each image's RGB intensity levels are analyzed, confirming their distribution predominantly within the range of 140 to 220. This consistency in intensity levels across parasitized and non-parasitized cells aids in a more straightforward analysis. Figures 4 and 5 provide an example of RGB color distribution and histogram.

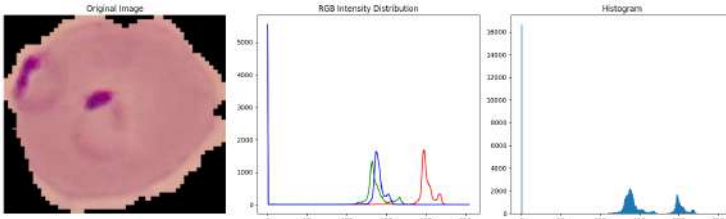


Figure 4. RGB Color Distribution and histogram in a sample parasitized image

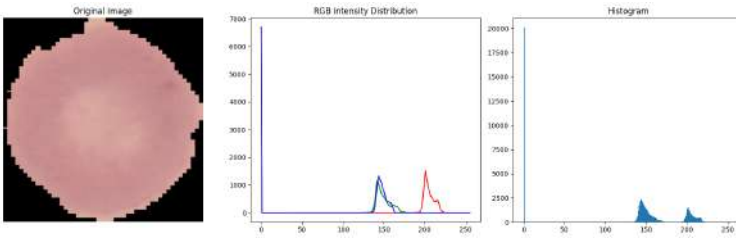


Figure 5. RGB Color Distribution and histogram in a sample non-parasitized image

- **Image Resizing and Normalization**

To accommodate the uniformity required for deep learning algorithms, all images in the study have been resized to a standard dimension of (224 224) pixels. Furthermore, standardization has been meticulously applied to each image, normalizing the pixel values to center around the mean, which is instrumental in expediting the convergence during the model training phase.

The formula for pixel normalization encapsulates this process is given by equation 1.

$$I' = \frac{I}{255} \quad (1)$$

where I is the original image matrix, and I' is the normalized image matrix.

- **Filtering - Noise Removal**

Images often capture extraneous variations attributed to factors like inconsistent lighting, surrounding environmental conditions, or particulate matter within the camera lens. Such variations, broadly called noise, can obscure image details and affect image quality (Figure 6). Since noise can adversely affect the performance of classification models, it is crucial to cleanse the images of these unwanted artefacts to ensure precise classifications [11].

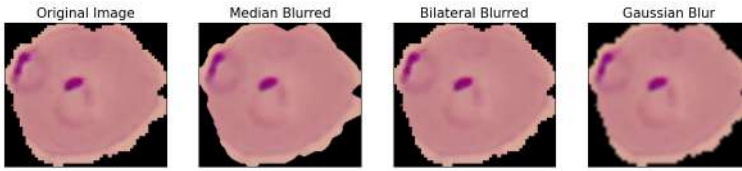


Figure 6. Noise Reduction Techniques in Image Preprocessing

Analysis of the depicted figures indicates the efficacy of Bilateral filtering in maintaining edge definition while reducing noise. Nonetheless, reviews of current literature suggest that median filtering stands out for its precision in noise reduction [6].

- **Data Augmentation**

Data augmentation techniques were employed to enhance the diversity of the dataset and simulate various imaging conditions. This process enriches the training dataset by generating altered versions of the images, thereby providing a more comprehensive set of patterns for the model to learn from. Utilizing the `'ImageDataGenerator'` function from *Keras*, we implemented a series of transformations, including random rotations (`'rotation_range=20'`), width and height shifts (`'width_shift_range=0.1'`, `'height_shift_range=0.1'`), shear intensity adjustments (`'shear_range=0.1'`), zooming actions (`'zoom_range=0.1'`), and horizontal flips (`'horizontal_flip=True'`). These transformations mimic variations due to patient movements and different slide preparations and help prevent overfitting, making the model more robust to unseen data.

3.3. Machine Learning Methods

All the model experiments have been implemented using the hyperparameters and callbacks mentioned in Table 1.

Table 1. Model Training Parameters

Hyperparameter	Value
Activation Function	ReLU, Sigmoid
Cost Function	Binary Cross-Entropy
Learning Rate	1 x 10 ⁻⁴
Optimizer	RMSprop
Epochs	20
Dropout Ratio	0.5
Batch Size	10
Training Callbacks	TensorBoard
	EarlyStopping
	ModelCheckpoint

• Convolutional Neural Networks (CNN)

The efficacy of CNN in extracting and processing features from raw image data forms the foundation of our system design. CNNs leverage the spatial hierarchies of pixels within an image, employing local receptive fields, shared weights, and pooling to reduce dimensionality while preserving essential features. This deep learning methodology is particularly suited to differentiating between parasitized and non-parasitized blood cells, which is critical for accurately diagnosing malaria.

The CNN architecture is shown in Figure 7, which starts with an input layer designed to handle images that are 224x224x3, which represent height, width, and RGB color channels. Next comes a series of convolutional layers that extract basic features using 32 filters with a 3x3 kernel size. These are paired with max-pooling layers that reduce the spatial dimensions by half, providing translational invariance while reducing computational complexity.

The architecture continues interleaving convolutions and pooling, with each stage incrementally abstracting higher-level features. Notably, the final convolutional stratum augments the filter quantity to 64, supplying an enriched feature set for subsequent classification. These features are then flattened into a one-dimensional array, facilitating the transition to the densely connected layers. A dense layer with 1024 neurons functions as an advanced feature classifier, followed by a dropout layer aimed at preventing overfitting by randomly disregarding neuron connections during training. The network culminates with a sigmoid activation function in the

output layer, providing the input image's classification probability. The architecture boasts 12,884,993 trainable parameters, indicative of the model's complexity and capacity for feature extraction and classification. The model's significant trainable parameter count underscores its comprehensive adaptability, honed through backpropagation during training. With a storage requirement of approximately 49.15 MB, the model's voluminous structure is well-suited for malaria detection from cell images.

- **EfficientNet-B3 Model Architecture**

The EfficientNet-B3 model is an advanced CNN designed to optimize computing resources and accuracy. Our adaptation accepts 224x224 RGB images and incorporates a compound scaling method that uniformly scales the network's width, depth, and resolution. The model's architecture starts with a convolutional base that captures complex patterns through a sequence of convolution layers, leading to a dense feature representation. A flattening layer transitions the 2D feature maps into a 1D vector, followed by a dense layer with dropout for regularization to mitigate overfitting. The final layer, a dense output with a sigmoid function, yields the classification probability, indicating the presence or absence of malaria parasites. Leveraging pre-trained weights from ImageNet, the model is fine-tuned on our dataset with a reduced learning rate, allowing it to adapt to the specifics of medical imaging for malaria detection.

- **YOLOv11m Model Architecture**

YOLOv11m represents the medium variant of the YOLOv11 family, an advanced object detection model adapted here for malaria parasite classification. Figure 7 provides the evolution of YOLO algorithms over the years. YOLOv11 introduces architectural innovations, including an enhanced backbone for superior feature extraction, an advanced neck for improved feature aggregation, and an optimized head for precise localization and classification. These improvements enable YOLOv11 to handle challenges like varying object sizes and overlapping instances effectively. Benchmarking on diverse datasets shows YOLOv11 (Figure 8) outperforming predecessors in accuracy while maintaining low inference time and a compact model size of approximately 20.1 million parameters [29, 30]. In this study, we fine-tuned YOLOv11m on segmented red blood cell images for binary classification (parasitized vs. uninfected), leveraging its efficiency for potential real-time deployment in resource-limited settings.

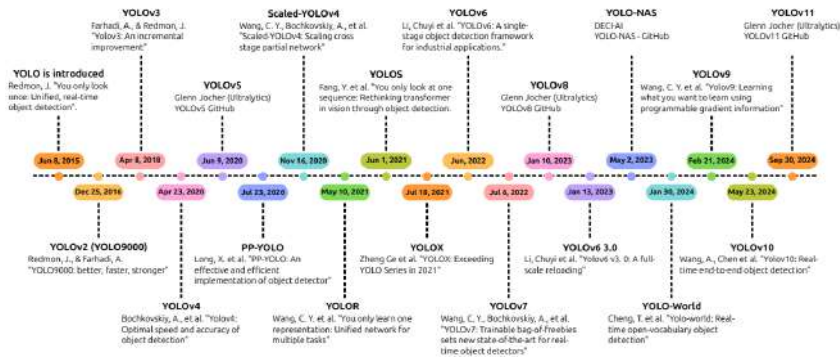


Figure 7. Evolution of YOLO Algorithms throughout the years [29].

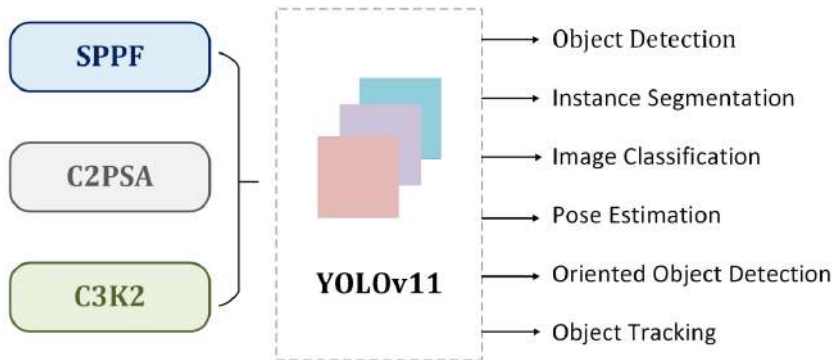


Figure 8. Key architectural modules in YOLO11 [30].

4. RESULTS AND DISCUSSIONS

4.1. Model Performance Evaluation Metrics

The performance of each model is evaluated using a suite of measurements, each reflecting different aspects of accuracy and reliability. The table 2 summarizing these metrics along with their respective equations and detailed descriptions [31].

Table 2. Evaluation metrics formulas with a detailed description.

Metric	Formulas	Description
Accuracy	$\frac{(TP + TN)}{(TP + TN + FP + FN)}$	Measures the overall correctness of the model by dividing the sum of correct predictions by the total number of predictions.
Precision	$\frac{TP}{(TP + FP)}$	Assesses the model's ability to return only relevant instances, showing the proportion of true positives among all optimistic predictions.
Recall	$\frac{TP}{(TP + FN)}$	Reflects the model's capability to identify all relevant instances, indicating the proportion of true positives detected from all actual positives.
F1 Score	$\frac{2 * (Precision * Recall)}{(Precision + Recall)}$	Balance precision and Recall are instrumental when the cost of false positives and false negatives differ significantly.
Matthews Correlation Coefficient (MCC)	$\frac{(TP * TN - FP * FN) / \sqrt{[(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)]}}$	Produces a high score only if the prediction obtained good results in all four confusion matrix categories relative to their sizes.
Area Under the ROC Curve (AUC ROC)	Integral of ROC curve	Represents the model's ability to discriminate between classes, with a higher area indicating better performance.
Area Under the Precision-Recall Curve (AUC PR)	Integral of Precision-Recall curve	Useful in imbalanced datasets, it shows the relationship between precision and recall for different probability thresholds.

4.2. Performance of the Models

We compared various deep-learning architectures to ascertain the most effective model for malaria detection from blood smear images. This section elucidates the comparative analysis, grounded on the performance during training, validation, and ultimately on unseen test data. The custom CNN model, designed as the foundational benchmark for this study, underwent 20 epochs of training, a duration sufficient to observe convergence trends without overfitting. In the comparative analysis of deep learning models for malaria detection, training accuracy is a crucial measure of a model's learning effectiveness over the training dataset. Figure 9 illustrates the evolution of training accuracy for six different models throughout the training epochs.



Figure 9. Training Accuracy Across Models

The CNN model shows notable learning efficiency, displaying trends similar to those of more sophisticated architectures, such as ResNet50 and VGG19. Notably, EfficientNet-B3 emerges as a frontrunner, reaching and sustaining peak accuracy levels more swiftly than its counterparts. This suggests that EfficientNet-B3's structure, which balances network depth, width, and resolution, may be particularly well-suited to this classification task. YOLOv11m demonstrates rapid convergence, achieving high accuracy with fewer parameters, making it suitable for efficient deployment. On the other end of the spectrum, the EfficientNet-B7 model lags slightly behind in training accuracy despite being part of the same family as EfficientNet-B3. This could be due to its complexity, which might require more epochs to fully capitalize on its learning capabilities, or it could be indicative of overfitting, where the model complexity does not translate to better performance on the given dataset. Figure 10 represents the training loss of the different models, illustrating the effectiveness of the learning process of each model over the epochs.

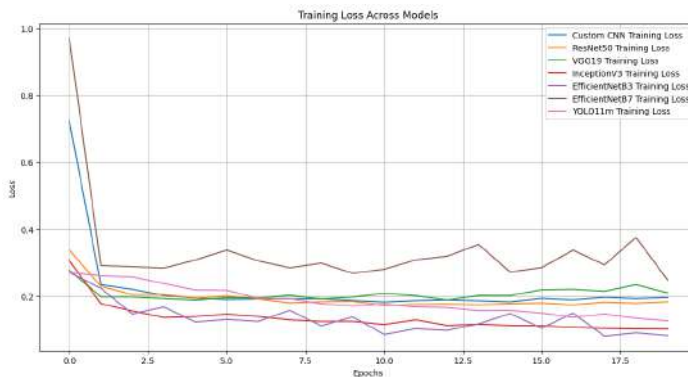


Figure 10. Training Loss Across Models

The Custom CNN model initiates its training with a substantial decrease in losses, reflecting a swift initial learning phase. On the other hand, the EfficientNet-B3 model demonstrates a consistent and uniform decline in loss, indicating a methodical and stable learning trajectory. The EfficientNet-B7, despite being from the same family, registers a higher loss comparatively. Other models, such as ResNet50 and VGG19, present an even reduction in loss, attaining a minimal and steady level early during training, which implies an effective learning mechanism with minimal danger of overfitting. The more complex InceptionV3 architecture also shows a reliable diminishment in loss across epochs. YOLOv11m exhibits low and stable loss, benefiting from its optimized architecture for faster training. The graph in Figure 11 illustrates the validation accuracy of various models throughout the training epochs.

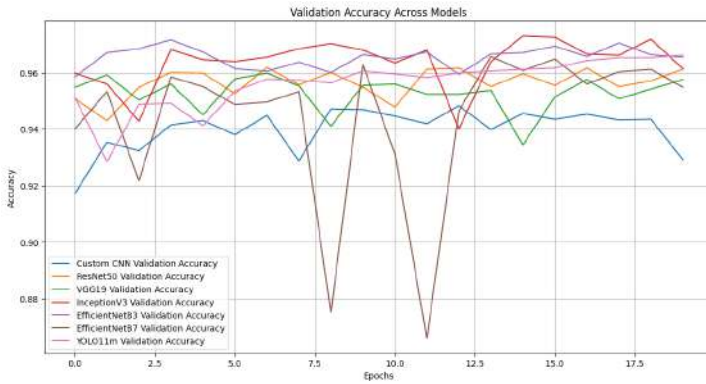


Figure 11. Validation Accuracy Across Models

Despite some observed fluctuations, the Custom CNN model demonstrates remarkable generalization ability, although slightly lower than the other models. The EfficientNet-B3 model is a consistent leader, consistently achieving the highest validation accuracy and demonstrating excellent predictive performance on previously unseen data. On the other hand, the EfficientNet-B7 model, despite its sophisticated structure, displays a lower generalization capacity than its EfficientNet equivalent and also shows fluctuations in validation accuracy. YOLOv11m maintains high validation accuracy with minimal fluctuations, underscoring its robustness.

4.3. Result on Test Data

Table 3. Summary of Test Evaluation Metrics Across Models (%).

Model	F1 Score	Accuracy	Precision	Recall	MCC	AUC ROC	AUC
Custom CNN	93.84	93.38	89.50	98.62	87.19	97.91	97.38
ResNet50	96.57	96.44	95.27	97.91	92.92	98.66	98.21
VGG19	96.13	95.99	94.89	97.41	92.01	98.69	98.35
InceptionV3	97.12	97.01	95.63	98.65	94.05	99.25	99.17
EfficientNet-B3	98.12	98.08	98.36	97.87	96.15	99.59	99.53
EfficientNet-B7	95.98	95.77	93.33	98.79	91.69	98.81	98.14
YOLOv11m	96.89	96.93	98.32	95.50	93.91	99.53	99.57

In summarizing the comparative performance of the deep learning models, Table 3 elucidates the remarkable proficiency of the EfficientNet-B3 model across all evaluated metrics. It achieves the highest F1 Score of 98.12%, Accuracy of 98.08%, and an impressive Matthews Correlation Coefficient (MCC) of 96.15%, underscoring its superior predictive capabilities. Precision stands at 98.36%, with Recall at 97.87%, highlighting the model's exceptional balance between sensitivity and specificity. YOLOv11m performs competitively, with a high precision of 98.32% and AUC PR of 99.57%, indicating its strength in minimizing false positives, which is critical for medical diagnostics.

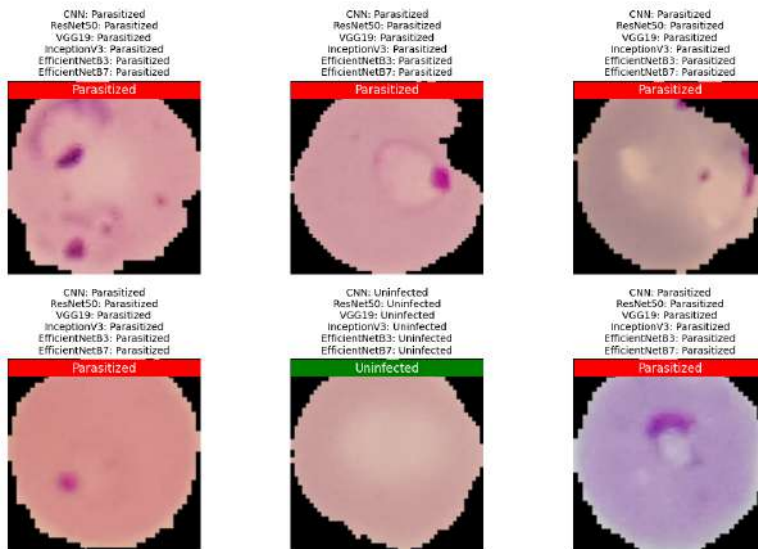


Figure 12. Classification Results from Different CNN Architectures

To complement the quantitative analysis visually, Figure 12 presents the classification results from the various CNN architectures, showcasing their predictive performance on sample images. This illustration reflects the numerical data from Table 5 and provides an intuitive understanding of how each model differentiates between ‘Parasitized’ and ‘Uninfected’ cells.

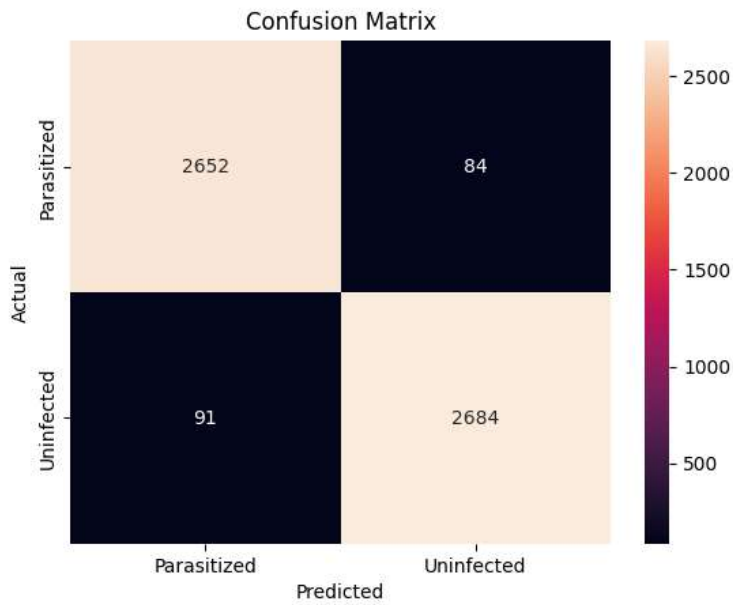


Figure 13. EfficientNet-B3 model Confusion Matrix

The confusion matrix for EfficientNet-B3 (Figure 13) further substantiates its efficacy, demonstrating a high actual positive rate with 2,652 parasitized cases accurately identified and 2,684 uninfected cases correctly recognized. The model’s precision is evidenced by its minimal misclassification, with only 84 parasitized cases mistaken as uninfected and 91 uninfected cases incorrectly labelled as parasitized. This minimal error margin underscores EfficientNet-B3’s potential as a leading model for reliable malaria detection, which is crucial for clinical settings where the stakes for accurate diagnosis are high.

To assess feasibility for real-time applications, we compared computational requirements across models (Table 4). YOLOv11m, with 20.1 million parameters, offers a balance between performance and efficiency, making it ideal for deployment on mobile or edge devices in low-resource settings. In contrast, EfficientNet-B7’s higher parameter count (192.24 million) may limit its practicality despite strong metrics.

Table 4. Computational Requirements for All Models (Parameters in Millions)

Model	Parameters (M)
Custom CNN	12.88
ResNet50	126.30
VGG19	45.72
InceptionV3	74.20
EfficientNet-B3	87.77
EfficientNet-B7	192.24
YOLOv11m	20.10

The high performance of EfficientNet-B3 and the lightweight YOLOv11m suggests their potential for deployment in clinical settings, particularly in malaria-endemic regions. YOLOv11m’s low computational requirements make it ideal for mobile or edge devices in low-resource settings, enabling rapid malaria diagnosis where access to advanced hardware is limited. Integration with existing healthcare systems could involve deploying these models on cloud platforms or portable devices, enhancing accessibility in rural areas.

Regarding false positives (FP) and false negatives (FN), EfficientNet-B3’s high precision (98.36%) minimizes FP, reducing unnecessary treatments and associated costs/side effects. Its recall (97.87%) limits FN, ensuring few infected cases are missed, which is vital to prevent disease progression and transmission. YOLOv11m’s even higher precision (98.32%) further reduces FP, though slightly lower recall (95.50%) indicates a trade-off; overall, both models mitigate diagnostic errors effectively compared to manual microscopy.

The implications of these findings are profound, especially considering the high burden of malaria and the limitations of traditional diagnostic methods. By automating the detection and classification process, deep learning models can assist in alleviating the workload on healthcare systems and improving the speed and accuracy of diagnoses, thus saving lives through timely treatment. However, this study has limitations. The reliance on a single dataset for model training and evaluation, while providing a controlled environment for comparison, does not account for the variability and complexity of real-world scenarios. To address this, future work will validate the models on additional datasets, such as the Malaria Bounding Box dataset, to ensure robustness across diverse imaging conditions. Future research should focus on applying these models to diverse and more extensive datasets, real-time analysis, and integrating such models into existing healthcare infrastructure. Furthermore, while the study focuses on binary classification, the natural progression of this work would involve multi-class classification to iden-

tify different life stages of malaria parasites, which is crucial for understanding the disease progression and tailoring treatment strategies accordingly. A comparison with other studies in the literature is given in Table 5.

Table 5. Comparison with other studies in the literature

Reference	Algorithms	F1 Score	Accuracy	Precision	Recall	AUC ROC
Maqsood et al. [11]	CNN	96.82	96.82	96.82	96.33	-
Nakasi et al. [18]	Faster-RCNN	-	93.03	72.29	-	-
Zamora et al. [32]	SVM	-	-	-	94.00	-
Bibin et al. [33]	DBN	-	-		97.60	-
Sarkar et al. [34]	VGG19	96.16	96.15	97.54	94.82	-
Quan et al. [19]	ADCN	97.50	97.47	-	97.86	-
Our work	EfficientNet-B3	98.12	98.08	98.36	97.87	99.59

5. CONCLUSION

This study has comprehensively evaluated a custom-designed Convolutional Neural Network (CNN) against five state-of-the-art transfer learning models in classifying malaria parasites from segmented red blood cell images. The results undeniably highlight the potential of deep learning techniques in public health, particularly for malaria-endemic regions in sub-Saharan Africa. The EfficientNet-B3 model’s performance was notably superior, with high scores in accuracy, precision, Recall, and other metrics, demonstrating its robustness in malaria parasite classification tasks. This underscores the effectiveness of CNNs, especially those with compound scaling and advanced optimization, for medical image analysis. YOLOv11m provided a compelling alternative, offering high precision and efficiency with fewer parameters, making it suitable for low-resource environments. While not surpassing the transfer learning models, the custom CNN still performed admirably, validating the potential for developing tailored deep learning solutions that are more accessible and potentially more adaptable to specific local contexts.

Author Contribution Rates

Design of Study: AA(%50), AM(%50)

Data Acquisition: AA(%50), AM(%50)

Data Analysis: AA(%50), AM(%50)

Writing Up: AA(%50), AM(%50)

Submission and Revision: AA(%50), AM(%50)

REFERENCES

- [1] C. Prevention, "for DC and. CDC-Malaria-About Malaria-Biology," *CDC-Centers Dis Control Prev*, 2020.
- [2] D. Posfai *et al.*, "Plasmodium parasite exploits host aquaporin-3 during liver stage malaria infection," *PLoS Pathog*, vol. 14, no. 5, p. e1007057, May 2018, doi: 10.1371/journal.ppat.1007057.10.1371/journal.ppat.1007057
- [3] E. Worrall, S. Basu, and K. Hanson, "Is malaria a disease of poverty? A review of the literature," *Tropical Medicine & International Health*, vol. 10, no. 10, pp. 1047-1059, 2005.
- [4] W. H. Organization, *World malaria report 2023*. World Health Organization, 2023.
- [5] T. Shavlakadze *et al.*, "Short-term Low-Dose mTORC1 Inhibition in Aged Rats Counter-Regulates Age-Related Gene Changes and Blocks Age-Related Kidney Pathology," *J Gerontol A Biol Sci Med Sci*, vol. 73, no. 7, pp. 845-852, Jun 14 2018, doi: 10.1093/gerona/glx249.10.1093/gerona/glx249
- [6] S. Mishra, "Malaria parasite detection using efficient neural ensembles," *Journal of Electronics, Electro-medical Engineering, and Medical Informatics*, vol. 3, no. 3, pp. 119-133, 2021.
- [7] J. P. Dailly, A. Minuti, and N. Khan, "Diagnosis, Treatment, and Prevention of Malaria in the US: A Review," *JAMA*, vol. 328, no. 5, pp. 460-471, Aug 2 2022, doi: 10.1001/jama.2022.12366.10.1001/jama.2022.12366
- [8] S. Rajaraman *et al.*, "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, p. e4568, 2018.
- [9] S. Savkare and S. Narote, "Automated system for malaria parasite identification," in *2015 international conference on communication, information & computing technology (ICCICT)*, 2015: IEEE, pp. 1-4.
- [10] V. K. Bairagi and K. C. Charpe, "Comparison of Texture Features Used for Classification of Life Stages of Malaria Parasite," *Int J Biomed Imaging*, vol. 2016, p. 7214156, 2016, doi: 10.1155/2016/7214156.10.1155/2016/7214156
- [11] A. Maqsood, M. S. Farid, M. H. Khan, and M. Grzegorzczek, "Deep malaria parasite detection in thin blood smear microscopic images," *Applied Sciences*, vol. 11, no. 5, p. 2284, 2021.
- [12] F. Sherif and A. Mohammed, "Detection of Malaria Infection Using Convolutional Neural Networks," in *2023 Intelligent Methods, Systems, and Applications (IMSA)*, 2023: IEEE, pp. 291-296.
- [13] E. Irmak, "A Novel Implementation of Deep-Learning Approach on Malaria Parasite Detection from Thin Blood Cell Images," (in English), *Electrica*, vol. 21, no. 2, pp. 216-224, May 2021, doi: 10.5152/electrica.2020.21004.10.5152/electrica.2020.21004
- [14] P. Krishnadas and N. Sampathila, "Automated detection of malaria implemented by deep learning in Py-Torch," in *2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECTT)*, 2021: IEEE, pp. 01-05.
- [15] M. O. F. Goni *et al.*, "Diagnosis of Malaria Using Double Hidden Layer Extreme Learning Machine Algorithm With CNN Feature Extraction and Parasite Inflator," *IEEE Access*, vol. 11, pp. 4117-4130, 2023.
- [16] B. Almuahaya, R. Mohammed, A. Mohammed, and B. Saha, "Malaria Detection Using Convolutional Neural Networks: A Comparative Study," in *2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, 2023: IEEE, pp. 1-8.
- [17] A. Agrawal, C. Hsu, and S. Tan, "A Comparative Analysis of Supervised and Semi-Supervised Learning Models for Malaria Classification Using Explainable AI Techniques."
- [18] R. Nakasi, E. Mwebaze, A. Zawedde, J. Tusubira, B. Akera, and G. Maiga, "A new approach for microscopic diagnosis of malaria parasites in thick blood smears using pre-trained deep learning models," *SN Applied Sciences*, vol. 2, pp. 1-7, 2020.

-
- [19] Q. Quan, J. Wang, and L. Liu, "An Effective Convolutional Neural Network for Classifying Red Blood Cells in Malaria Diseases," *Interdiscip Sci*, vol. 12, no. 2, pp. 217-225, Jun 2020, doi: 10.1007/s12539-020-00367-7.
 - [20] V. Magotra and M. K. Rohil, "Malaria Diagnosis Using a Lightweight Deep Convolutional Neural Network," *Int J Telemed Appl*, vol. 2022, p. 4176982, 2022, doi: 10.1155/2022/4176982.
 - [21] K. M. Pimple, P. P. Likhitar, and S. Pande, "Convolutional neural networks for malaria image classification," in *Proceedings of Data Analytics and Management: ICDAM 2021, Volume 2*, 2022: Springer, pp. 459-470.
 - [22] Z. Liang et al., "CNN-based image analysis for malaria diagnosis," in *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*, 2016: IEEE, pp. 493-496.
 - [23] B. Lufyaila, B. Mgawe, and A. Sam, "Fine-Tuned YOLO-Based Deep Learning Model for Detecting Malaria Parasites and Leukocytes in Thick Smear Images: A Tanzanian Case Study," *Machine Learning with Applications*, p. 100687, 2025.
 - [24] *Google Colab*. [Online]. Available: <https://colab.research.google.com>
 - [25] *Keras*. [Online]. Available: <https://keras.io/>
 - [26] *TensorFlow*. [Online]. Available: <https://www.tensorflow.org>
 - [27] *Ultralytics YOLO11*. [Online]. Available: <https://docs.ultralytics.com/fr/models/yolo11>
 - [28] *Kaggle*. [Online]. Available: <https://www.kaggle.com/>
 - [29] N. Jegham, C. Y. Koh, M. Abdelatti, and A. Hendawi, "Yolo evolution: A comprehensive benchmark and architectural review of yolov12, yolo11, and their previous versions," *arXiv preprint arXiv:2411.00201*, 2024.
 - [30] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, 2024.
 - [31] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation metrics in machine learning algorithms," in *Computer Science On-line Conference, 2023*: Springer, pp. 15-25.
 - [32] A. Pages-Zamora, M. Cabrera-Bean, and C. Diaz-Vilor, "Unsupervised online clustering and detection algorithms using crowdsourced data for malaria diagnosis," *Pattern Recognition*, vol. 86, pp. 209-223, 2019.
 - [33] D. Bibin, M. S. Nair, and P. Punitha, "Malaria parasite detection from peripheral blood smear images using deep belief networks," *IEEE Access*, vol. 5, pp. 9099-9108, 2017.
 - [34] S. Sarkar, R. Sharma, and K. Shah, "Malaria detection from RBC images using shallow convolutional neural networks," *arXiv preprint arXiv:2010.11521*, 2020.
-