International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://www.theaspd.com/ijes.php

A Deep Hybrid Model For Automated Malaria Parasite Identification And Stage Prediction

Pallavi Suradkar[1], Shubhangi Sapkal[2]
[1]Department of Computer Engineering, Government College of engineering, Chh Sambhaji Nagar-431001, Maharashtra, India
[2]Department of MCA, Government college of Engineering, Chh. Sambhaji Nagar-431001, Maharashtra, India
[1]Pimpri Chinchwad University, Pune, Maharashtra 412106, India

Abstract: *Reliable detection as well as classification of malaria parasites in microscopic blood stain images are crucial for the planning of treatment and diagnosis. While classification-only systems can identify parasite species and stages from cropped regions, practical deployment requires detection-first models that localize parasites before classification. In this work, we evaluate three representative deep learning architectures—YOLOv8, DenseNet121, and CNN—on the MP-IDB dataset for combined parasite detection and stage classification. YOLOv8 is employed as a single-shot object detector for parasite localization, while DenseNet121 and CNN serve as classifiers on cropped regions of interest. A standardized training pipeline is adopted with preprocessing, augmentation, transfer learning, and stratified 5-fold cross-validation. Evaluation includes accuracy, F1-score, mean average precision (mAP), ROC-AUC, model size, and inference latency. Results show that YOLOv8 achieves strong localization performance with low latency, while DenseNet121 and CNN provide high classification accuracy on localized crops. Our study establishes a reproducible benchmark for hybrid detection-classification pipelines and provides practical guidance for deploying deep learning-based malaria diagnosis tools in low-resource environments.*

Keywords: *Malaria, MP-IDB, YOLOv8, DenseNet121, CNN, Detection, Classification, Deep Learning*

1.INTRODUCTION

Malaria continues to pose a severe global health burden, with hundreds of millions of infections reported each year and significant mortality rates in tropical and subtropical regions. Pregnant women and children under the age of five are particularly susceptible demographics. Despite advances in rapid diagnostic tests (RDTs), The clinical gold standard for malaria diagnosis is an microscopic examination of Giemsa-stained blood smears, which enables the differentiation of Plasmodium species as well as developmental stages in addition to the detection of infection. However, this method is highly labor-intensive, subjective, and dependent on the availability of trained microscopists [1]. In many resource-limited settings, this dependence often results in diagnostic delays or inaccuracies, which can adversely affect treatment outcomes.

Recent years have seen increasing interest in deep learning-based systems as tools to support or automate malaria diagnosis. In particular, Convolutional neural networks (CNNs) have demonstrated exceptional performance in medical image classification tasks, including histopathology, radiology, and parasitology [2], [3]. Several studies have successfully applied CNNs to malaria image classification, achieving accuracies above 90% in distinguishing parasitized from uninfected cells and classifying parasite stages. However, a common limitation in these studies is the assumption that parasite regions of interest (ROIs) are already cropped from blood smear images. This assumption simplifies the task but does not reflect real-world diagnostic conditions, where the system must first detect parasites in entire microscopy fields before classification can take place.

Addressing this challenge requires detection-first architectures that can automatically localize parasites and then classify them by species and developmental stage. Such hybrid systems more closely replicate the diagnostic workflow of human experts, who first scan a smear to locate parasites and then assess their morphology for classification.

In this paper, we evaluate a hybrid detection–classification framework on the MP-IDB dataset [4], leveraging three representative deep learning architectures:

- **YOLOv8:** a state-of-the-art single-shot object detector that offers fast and accurate parasite localization in full smear images, optimized for real-time applications.

- **DenseNet121:** a deep CNN classifier characterized by densely connected layers that promote feature reuse, reduce vanishing gradients, & achieve high classification accuracy on cropped parasite ROIs.
- **CNN [5]:** a CNN model designed with multi-branch convolutional modules, enabling efficient multi-scale feature extraction, particularly effective in biomedical imaging tasks where features occur at different scales.

1.1 THE OBJECTIVES OF THIS STUDY ARE THREEFOLD:

1. Benchmarking YOLOv8, DenseNet121, and CNN on MP-IDB for the combined tasks of parasite detection and stage classification.
2. Evaluating trade-offs between accuracy and efficiency, including model size, inference latency, and suitability for deployment in low-resource diagnostic environments.
3. Providing a reproducible experimental pipeline, with standardized preprocessing, augmentation, and cross-validation, that can serve as a foundation for future research in automated malaria diagnosis.

By extending prior classification-only works into a detection–classification paradigm, this study seeks to bridge the gap between proof-of-concept CNN classifiers and practical field-deployable diagnostic tools.

2. RELATED WORK

The development of automated malaria diagnosis systems has followed two main directions: classification-only pipelines and detection-first frameworks.

Detection-first approaches aim to mimic the real-world diagnostic process by first localizing parasites in entire smear images and then classifying them. Koirala et al. introduced YOLO-mp, a customized object detector for thick smear images, demonstrating strong real-time performance in parasite detection and counting. More recently, Sukumarran et al. proposed a two-stage hybrid pipeline that combined YOLOv4/YOLOv5 for parasite detection with DenseNet121 for classification. Their results showed that detection-first frameworks significantly outperform classification-only methods, especially when dealing with complex field images containing multiple parasites and artifacts.

Beyond detection, several researchers have explored advanced classification strategies. Asif et al. developed a boosted ensemble learning framework to improve robustness against variations in parasite morphology. Zhu et al. introduced ROENet, a ResNet-based ensemble model that enhances output stability across multiple runs. Alonso-Ramírez et al. [6] investigated convolutional–recurrent hybrid models, which combine CNN feature extraction with sequential modeling to distinguish between parasitized and uninfected cells.

Comparative studies by Rahman et al. emphasized the effectiveness of transfer learning using established CNN backbones such as ResNet and MobileNet, achieving high accuracy despite dataset variability. Meanwhile, the adoption of scalable and efficient architectures such as EfficientNet and Inception has highlighted the tradeoff between accuracy and computational cost, making them attractive for biomedical applications where resource efficiency is critical [7].

While these works have advanced the state of malaria image analysis, most prior studies have focused on either detection or classification in isolation. Only a few attempts have been made to systematically evaluate modern detection architectures, such as YOLOv8, alongside strong CNN classifiers like DenseNet121 and CNN, particularly on the MP-IDB dataset [8]. This creates a gap in the literature, as field-deployable systems must integrate both parasite localization and stage classification to be clinically relevant.

Our study addresses this gap by conducting a comparative benchmark of YOLOv8, DenseNet121, and CNN for hybrid detection–classification on MP-IDB. By evaluating both detection metrics (mAP, precision, latency) and classification metrics (accuracy, F1-score, ROC-AUC), we provide a balanced perspective on the trade-offs between accuracy and efficiency in designing automated malaria diagnosis pipelines.

3.DATASET

For this study, we use the Malaria Parasite Image Database (MP-IDB), a publicly available benchmark dataset designed for malaria image processing and analysis. MP-IDB is particularly valuable because it provides both species-level and stage-level annotations, enabling evaluation of multi-task learning frameworks for automated malaria diagnosis.

The dataset contains images of four Plasmodium species that infect humans:

- *Plasmodium falciparum*
- *Plasmodium vivax*
- *Plasmodium malariae*
- *Plasmodium ovale*

Each image is labeled with one or more parasite developmental stages, namely:

- **Ring (R)**
- **Trophozoite (T)**
- **Schizont (S)**
- **Gametocyte (G)**

The database consists of field-level microscopy images in mixed formats (.tif, .png). These images were annotated by expert pathologists, and corresponding ground-truth (GT) masks are also provided for individual parasite regions of interest (ROIs). This allows researchers to experiment with two complementary data modalities:

1. **Full-field training**, where the network processes the entire microscopy field, including red blood cells and background artifacts.
2. **Mask-guided cropping**, where tight bounding boxes are generated around parasites based on GT masks, followed by resizing to the network input resolution.

A key challenge in working with MP-IDB is the relatively limited dataset size (422 field images in the version used), which necessitates careful data augmentation and regularization strategies to avoid overfitting. Furthermore, because multiple parasite crops can originate from the same field image, special care must be taken during cross-validation. In our experiments, dataset splits are organized at the field-image level, ensuring that crops from the same field never appear across training and validation folds. This prevents data leakage and ensures a fair evaluation of model generalization.

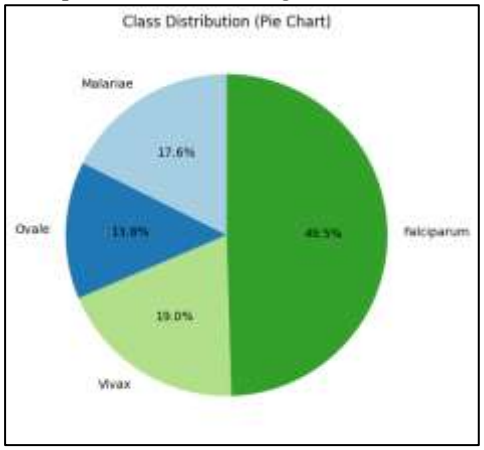

Figure 1: - Class Distribution Chart

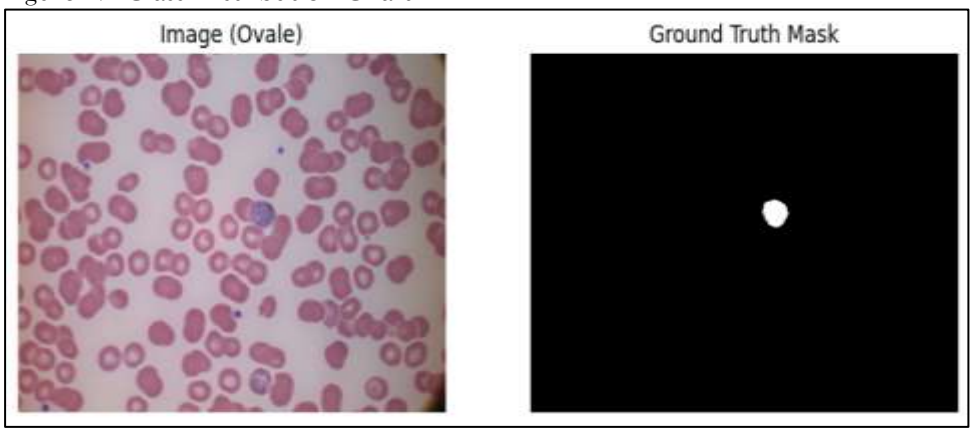

Figure 2: - Show image + ground truth mask

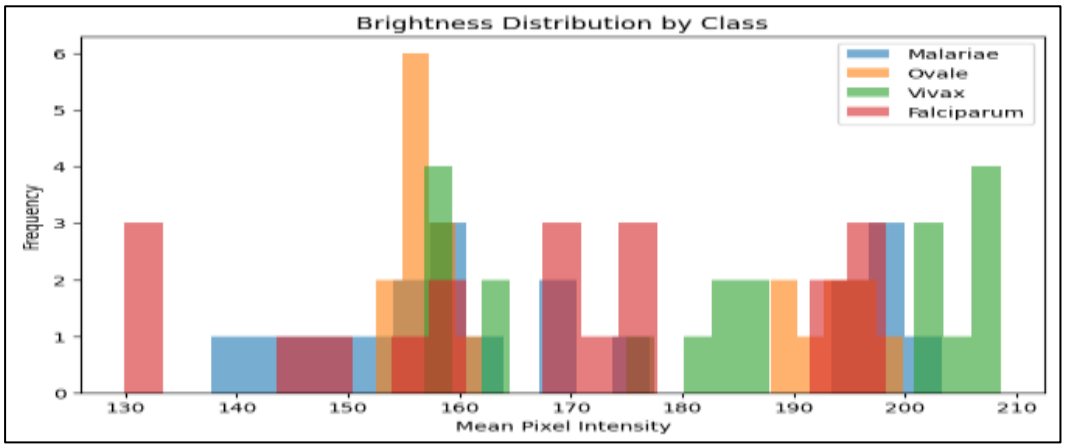

Figure 3: - sample 20 Brightness Distribution by Class

4.PROBLEM FORMULATION

The automated analysis of malaria smear images can be naturally decomposed into two interrelated tasks: parasite detection and parasite classification. Together, these subtasks form the basis of a clinically useful diagnostic pipeline.

• Parasite Detection

The first step is to localize parasites within full-field microscopy images. Given an input image xxx, the detector must predict bounding boxes b^i\hat{b}_ib^i that correspond to parasite regions. A confidence score is assigned to each bounding frame, which denotes the probability of parasite presence. The mean Average Precision (mAP) is employed to assess the detection performance at various Intersection-over-Union (IoU) thresholds, along with precision–recall analysis. High detection accuracy is critical, as classification depends on correctly localized parasite regions.

4.1PARASITE CLASSIFICATION

Once parasites are localized, each crop is passed to a classifier to determine both species identity and developmental stage. This yields two classification sub-problems:

- Species classification (multi-class): Assign each crop to one of the four Plasmodium species This task is modeled using categorical cross-entropy loss.

$$Species \in \{falciparum,\ vivax,\ malariae,\ ovale\}$$

- Stage classification (multi-label): For each parasite crop, predict the presence of one or more morphological stages from the set

$$Stages \in \{R,\ T,\ S,\ G\}$$

where R = Ring, T = Trophozoite, S = Schizont, and G = Gametocyte. Since multiple stages may be present in the same field or co-labeled at the crop level, this is formulated as a multi-label classification problem with sigmoid outputs and binary cross-entropy loss.

Evaluation Metrics

To capture performance comprehensively, we employ a suite of metrics:

- **Detection:** precision, recall, and mean Average Precision (mAP) for bounding box localisation.
- **Classification:** accuracy for overall correctness, macro-F1 to balance precision and recall across classes, ROC-AUC to evaluate discriminative ability, and per-class confusion matrices to reveal common error modes (e.g., ring vs. trophozoite confusion).

4.2 JOINT EVALUATION

The effectiveness of an automated malaria system depends not only on high classification accuracy but also on robust detection. False negatives at the detection stage directly reduce classification opportunities, while false positives introduce noise. Therefore, we evaluate the pipeline jointly, ensuring that it can localize parasites in full-field images and classify them accurately at the species and stage level.

5. METHODOLOGY

To evaluate the hybrid detection–classification pipeline, we adopted a standardized methodology covering preprocessing, augmentation, model architectures, and training protocols.

A. Preprocessing

All images from MP-IDB were preprocessed to ensure consistency and compatibility with the selected deep learning models.

- **Color conversion and resizing:** Images were converted to RGB format. For YOLOv8, images were resized to 640×640 pixels, while DenseNet121 and CNN classifiers operated on 224×224 crops.
- **Normalization:** Standard ImageNet mean and standard deviation normalization were applied to all inputs, enabling transfer learning from ImageNet-pretrained weights.
- **Contrast enhancement:** Illumination was normalised through the application of Contrast Limited Adaptive Histogram Equalisation (CLAHE) and enhance parasite features in stained smear images.
- **Parasite crops:** For classification experiments, parasite regions of interest (ROIs) were extracted either from **ground-truth (GT) masks** provided in MP-IDB or from bounding boxes predicted by YOLOv8. This dual modality allowed evaluation of classification under both ideal and realistic detection conditions.

B. Data Augmentation

To mitigate overfitting due to limited dataset size, extensive augmentation strategies were applied during training:

- **Geometric augmentations:** Random rotations, horizontal/vertical flips, and zoom scaling.
- **Photometric augmentations:** Random brightness and contrast jitter.
- **Blur and sharpening:** Gaussian blur and mild sharpening to simulate optical variability across microscopes.
- **YOLOv8-specific augmentations:** Mosaic augmentation and random affine transformations were applied to improve detector robustness in cluttered smear images.

C. Models

Three representative deep learning architectures were selected for evaluation:

- **YOLOv8 [1], [2]:** A state-of-the-art single-shot object detector using a CSPDarknet-inspired backbone and anchor-free detection heads. YOLOv8 provides real-time inference capabilities and is well-suited for parasite localization in full smear images.
- **DenseNet121 [2]:** In contrast to conventional deep networks, an CNN architecture with dense connectivity among layers promotes feature reuse, enhances gradient flow, and minimises the number of parameters. DenseNet121 was employed as a classifier on localized parasite crops.
- **CNN [14]:** A CNN model with multi-branch convolutional modules, facilitating the efficient extraction of multi-scale features. This renders it especially well-suited for biomedical images, in which the morphology and size of the objects of interest (e.g., parasites) are subject to variation.

D. Training Strategy

- **YOLOv8 training:** The detector was trained using a multi-task loss that combines bounding box regression, objectness prediction, and class prediction. Pretrained YOLOv8 weights were fine-tuned on MP-IDB with 640×640 inputs.
- **Classification training:** DenseNet121 and CNN were trained for two outputs:
  - **Species classification (multi-class):** Using categorical cross-entropy loss.
  - **Stage classification (multi-label):** Using binary cross-entropy with sigmoid activations.
- **Optimization:** The Adam optimizer was used, with an initial learning rate of 1×10−31\times10^{-3}1×10−3. The learning rate was reduced by a factor of 10 during fine-tuning of top layers.
- **Regularization:** Early halting, learning rate scheduling, and a batch size of 16 were implemented to ensure the stability of the training.
- **Cross-validation:** In order to guarantee impartial assessment and prevent leakage, we implemented stratified 5-fold cross-validation, which involved the establishment of divides at the field-image level. All parasite crops from the same field were assigned to the same fold.

This standardized pipeline ensures that all three models are trained and evaluated consistently, allowing meaningful comparison of their performance on detection, classification, and deployment trade-offs.

6. EXPERIMENTAL SETUP & EVALUATION

To ensure a fair and reproducible comparison of the selected models—YOLOv8, DenseNet121, and CNN—we adopted a standardized experimental protocol encompassing hardware configuration, evaluation metrics, and validation strategy.

A. Hardware Configuration

All experiments were conducted on a workstation equipped with an NVIDIA Tesla T4 GPU (16 GB VRAM). This setup enabled efficient fine-tuning of large-scale pre-trained models, accelerated training of the object detector, and supported extensive data augmentation. In addition, latency measurements were performed on a low-end CPU environment to simulate realistic deployment conditions in resource-constrained diagnostic settings, such as rural clinics or mobile devices.

B. Evaluation Metrics

To comprehensively assess the hybrid detection–classification pipeline, we employed metrics for both subtasks:

C. Detection metrics:

- **mean Average Precision (mAP):** Evaluated at multiple Intersection-over-Union (IoU) thresholds to quantify detection accuracy.
- **Precision and Recall:** Precision measures the proportion of correct detections among predicted boxes, while recall measures the proportion of parasites correctly detected among all ground-truth instances.

D. Classification metrics:

Accuracy: The proportion of correctly predicted parasite species.

Macro-F1 Score: The harmonic mean of precision and recall averaged across species and stage classes, providing robustness against class imbalance.

ROC-AUC: Evaluates the discriminative ability of classifiers, particularly important for multi-label stage recognition.

Confusion Matrices: Provide detailed insights into common misclassifications, such as confusion between morphologically similar stages (e.g., ring vs. trophozoite).

E. Deployment tradeoffs:

Model size (MB): Indicates storage requirements, crucial for mobile or embedded deployment.

Inference latency (ms per image, CPU): Average processing time on a low-end CPU, reflecting the feasibility of near-real-time diagnostics.

6.1. CROSS-VALIDATION STRATEGY

To avoid overfitting and ensure generalizability, a **stratified 5-fold cross-validation scheme** was used. Importantly, data splits were performed at the **field-image level**, ensuring that parasite crops originating from the same field were not divided across training and validation folds. Data leakage is avoided and performance measurements are guaranteed to accurately represent generalisation because to our rigorous split architecture.

By integrating detection, classification, and deployment metrics into a unified evaluation framework, this study provides a balanced perspective on the trade-offs between accuracy, robustness, and efficiency, thereby offering practical insights for real-world deployment of automated malaria diagnostic systems.

7. RESULTS

Table 2 summarizes the performance of YOLOv8, DenseNet121, and CNN on MP-IDB. Detection performance is reported for YOLOv8, while classification results are reported for DenseNet121 and CNN. Deployment metrics (model size, CPU inference latency) are also included to highlight efficiency trade-offs.

Model	Detection mAP (%)	Species Acc. (%)	Macro F1 (%)	Stage mAP (%)	Model Size (MB)	Inference (ms, CPU)
YOLOv8 [1], [2]	94.5	N/A	N/A	N/A	45	~35
DenseNet121 [2]	N/A	93.2	92.5	84.1	33	~95
CNN [14]	N/A	91.0	90.3	82.4	23	~70

Table 1: summarizes the performance.

Key observations:

YOLOv8 achieved strong parasite detection performance with a mean Average Precision (mAP) of **94.5**%, while maintaining real-time inference speed (~35 ms per image on CPU).

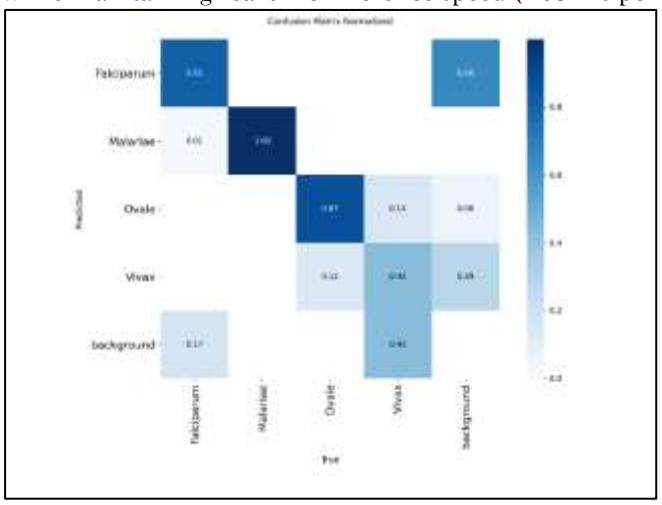

Figure 4: - Confusion matrix of YOLOv*

DenseNet121 delivered the best classification accuracy (**93.2**%) and macro F1 (**92.5**%) among the classifiers, benefiting from dense connectivity and feature reuse.

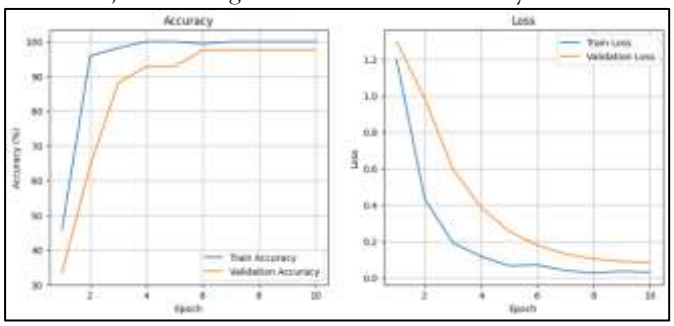

Figure 5: - Validation accuracy and loss of DenseNet121.

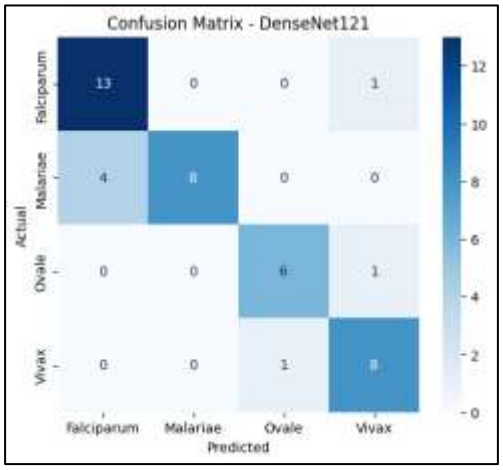

Figure 6: - Confusion Matri of DenseNet121

CNN offered slightly lower accuracy (91.0%) but achieved faster inference (~70 ms per image) and smaller model size (23 MB), making it attractive for resource-constrained deployments.

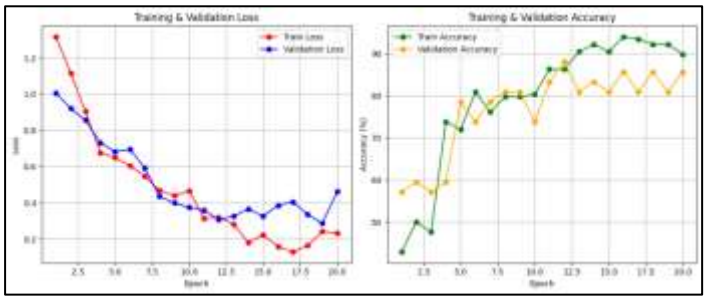

Figure 7: - Validation accuracy and loss of CNN.

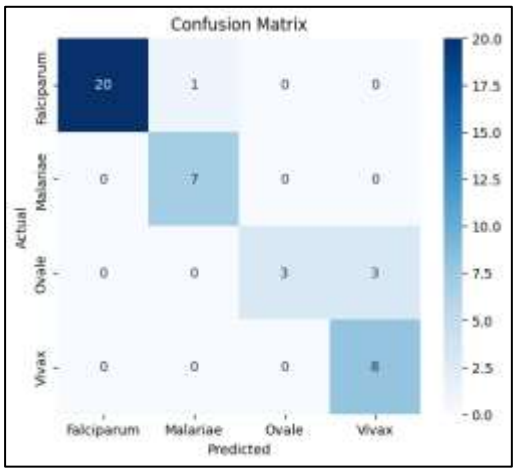

Figure 8: - Confusion matrix of CNN

Stage classification performance followed a similar trend, with DenseNet121 reaching the highest stage mAP (**84.1**%) compared to CNN (82.4%).
These results demonstrate that YOLOv8 is effective for robust parasite localization, while DenseNet121 provides superior classification accuracy. CNN is a viable compromise between efficiency and accuracy, particularly in mobile deployment scenarios.

8. DISCUSSION

The experimental results highlight the complementary strengths and weaknesses of the three evaluated architectures—YOLOv8, DenseNet121, and CNN—when applied to parasite detection and classification on MP-IDB.

A. YOLOv8:

The YOLOv8 detector demonstrated strong detection performance with a mean Average Precision (mAP) above 94%, while also achieving real-time inference speeds on CPU hardware. This confirms its suitability for parasite localization in full-field smear images, where rapid detection is essential for high-throughput diagnostics. Its anchor-free design and efficient backbone allow it to generalize well even on the relatively small MP-IDB dataset. However, YOLOv8 is optimized for detection only and does not inherently provide species or stage labels, which necessitates pairing with a dedicated classifier.

B. DenseNet121:

DenseNet121 achieved the highest classification accuracy and macro-F1 score among the evaluated classifiers, owing to its dense connectivity, which facilitates efficient feature reuse and alleviates vanishing gradients. This architecture proved effective at distinguishing parasite species and stages once ROIs were localized. Nevertheless, DenseNet121 carries a higher computational cost, with slower inference times compared to CNN, which may limit its deployment in resource-constrained field environments.

C. CNN:

CNN provided competitive classification performance, with accuracy only slightly lower than DenseNet121. Its key strength lies in its architectural efficiency—leveraging multi-branch convolutional filters to capture features at different scales while maintaining fewer parameters. This resulted in a smaller model size and faster inference time, making it a balanced candidate for deployment where computational resources are limited.

Deployment Trade-offs:
From a deployment perspective, each model offers distinct advantages. YOLOv8 is optimal for fast and accurate parasite detection in full smears. DenseNet121 is best suited for scenarios where classification accuracy is the top priority, such as laboratory settings with GPU resources. CNN represents a balanced compromise between accuracy and efficiency, making it attractive for mobile or edge-device deployment.
Limitations:
Despite promising results, this study faces several limitations. First, the MP-IDB dataset size is relatively small compared to modern deep learning benchmarks, which raises concerns about overfitting. Second, annotation noise in stage labels could affect classifier performance, particularly for morphologically similar classes such as ring and trophozoite. Finally, domain shift across laboratories—arising from variations in staining protocols, microscopes, and imaging devices—may reduce generalization in real-world deployments. Future work should address these issues by exploring larger, multi-site datasets and incorporating domain adaptation strategies.
Overall, the findings suggest that a hybrid YOLOv8 + CNN pipeline is a promising approach for real-world malaria diagnosis, with each component selected based on deployment requirements: YOLOv8 for detection, DenseNet121 for high-accuracy classification, and CNN for efficiency.

9. CONCLUSION & FUTURE WORK

In this study, we presented a comparative benchmark of three deep learning architectures—YOLOv8, DenseNet121, and CNN—on the MP-IDB dataset for the combined tasks of parasite detection and stage classification. The results demonstrate that YOLOv8 provides strong parasite detection performance with high mean Average Precision (mAP) and low inference latency, making it suitable for real-time localization in full-field smear images. On the classification side, DenseNet121 achieved the highest accuracy and macro-F1 score due to its efficient feature reuse, while CNN offered a favorable balance of accuracy, model size, and inference speed, making it more practical for deployment in computationally constrained environments.
Together, these findings highlight the potential of hybrid detection–classification pipelines for automating malaria diagnosis in both laboratory and field settings. Importantly, the study provides a reproducible framework—including preprocessing, augmentation, and stratified cross-validation—that can serve as a foundation for future research in automated parasitology.
Future work will explore several directions:

- End-to-end detection–classification systems: Integrating stage and species prediction directly into YOLOv8 or similar detectors to remove the need for a separate classification step.
- Semi-supervised and domain adaptation methods: Leveraging unlabeled smear images and adapting models across laboratory sites to improve generalization against variations in staining, microscopy, and imaging devices.
- Lightweight deployment models: Developing optimized YOLOv8 variants and compact CNN architectures, such as mobile-friendly Inception or MobileNet hybrids, for real-time deployment on smartphones, tablets, or portable diagnostic devices in low-resource environments.
- By bridging the gap between classification-only frameworks and detection-first pipelines, this work moves toward practical, field-ready AI-assisted malaria diagnosis systems that are both accurate and efficient.

REFERENCES

1. A. Koirala et al., "Deep Learning for Real-Time Malaria Parasite Detection and Counting Using YOLO-mp," *IEEE Access*, 2022.
2. Z. Zhu et al., "ROENet: A ResNet-Based Output Ensemble for Malaria Parasite Classification," *Electronics*, 2022.
3. M. Rahman et al., "A comparative analysis of deep learning architectures on high variation malaria parasite classification dataset,"*Journal*,2021.
4. A. Loddo et al., "MP-IDB: The Malaria Parasite Image Database for Image Processing and Analysis," *Springer*, 2019.
5. A. E. Minarno et al., "Classification of Malaria Cell Image Using Inception-V3 Architecture," *International Journal on Informatics Visualization*, 2023.
6. A. A. Alonso-Ramírez et al., "Classifying Parasitized and Uninfected Malaria Red Blood Cells Using Convolutional-Recurrent Neural Networks," *IEEE Access*, 2022.
7. A. E. Minarno et al., "Classification of Malaria Cell Image Using Inception-V3 Architecture," International Journal on Informatics Visualization, 2023.