Integrating Deep Learning for Malaria Cell Morphology Analysis and Subtype Classification

Aayush Sangani [1] **Antonela Tamagnini** [1]

Abstract

This study addresses the critical challenge of improving malaria diagnosis through advanced deep learning techniques. We propose an integrated approach combining object detection and instance segmentation to analyze malaria cell morphology and classify subtypes. Faster R-CNN was employed for initial cell localization and detection, while Mask R-CNN extended these capabilities by incorporating instance segmentation to generate bounding boxes and segmentation masks for enhanced feature extraction. MobileNet and VGG16 architectures were further utilized for multiclass classification of six cell types. The dataset, comprising infected and uninfected cell images, was preprocessed with normalization, augmentation, and class balancing techniques to enhance robustness. Results demonstrated 90.92% accuracy with MobileNet for classification tasks and improved cell detection and segmentation performance using Mask R-CNN. Compared to prior approaches limited to binary classification, our method achieves superior diagnostic precision and scalability, paving the way for future advancements in segmentation and real-time diagnostics.

1. Introduction

Problem before solution: Malaria remains a critical public health challenge, with effective diagnostic methods being essential for its control. Traditional microscopy-based methods for malaria diagnosis are time-intensive and require skilled personnel, making them unsuitable for resource-constrained settings. Automated solutions leveraging deep learning have shown promise, yet existing approaches primarily focus on binary classification, lacking detailed morphological insights necessary for advanced diagnostics.

Recent advancements in object detection, particularly the application of Faster R-CNN, have improved the detection of malaria-infected cells. However, these methods often fall short in addressing complex segmentation tasks that are vital for subtype classification and morphological analysis (Hung et al., 2017). To address these limitations, we explore an approach integrating object detection and instance segmentation to enhance diagnostic precision.

Research contributions: In this paper, we extend the capabilities of Faster R-CNN by incorporating Mask R-CNN to integrate instance segmentation with object detection. This enables simultaneous localization, classification, and mask generation for malaria cells. Further, we employ MobileNet and VGG16 for multiclass classification across six cell types. Our contributions include:

- A novel pipeline combining Mask R-CNN with multiclass classifiers for detailed malaria cell analysis.
- Preprocessing techniques, including normalization and augmentation, to improve model robustness.
- Comprehensive evaluation demonstrating a 90.92% classification accuracy and improved segmentation results compared to existing methods.

This work represents a significant step towards scalable, automated malaria diagnostics, predicting the stage of malaria in a patient, with potential for real-time application in resource-limited environments.

2. Related Work

Malaria diagnosis has long relied on manual microscopic analysis, a process that is time-consuming, error-prone, and dependent on skilled technicians. In recent years, advancements in computer vision and deep learning have sought to automate and enhance the accuracy of malaria detection and classification.

Hung et al. (2017) introduced the application of Faster R-CNN for object detection on malaria images, demonstrating the potential of deep learning models to localize

[1] Worcester Polytechnic Institute, Worcester, MA, USA.

CS/DS 541 Deep Learning Final Project, Worcester Polytechnic Institute.

and classify infected cells with significant accuracy (Hung et al., 2017). While their work emphasized the efficiency of Faster R-CNN for object detection tasks, it was limited in its ability to delineate cell boundaries and identify subcellular structures, which are crucial for subtype classification and morphological analysis.

Building on the limitations of previous binary classification models, another study explored malaria cell image classification using various convolutional neural networks (CNNs) like VGG-16 and ResNet-50 (Chima et al., 2020). The study focused on improving classification accuracy by leveraging transfer learning and data augmentation techniques. Despite achieving high classification performance, the approach lacked integration of instance segmentation, which is critical for understanding cell morphology and extracting fine-grained features.

Cheuque et al. (2022) presented an efficient multi-level convolutional neural network (CNN) approach for classifying white blood cells, offering insights into cell morphology by combining hierarchical feature extraction with robust classification (Cheuque et al., 2022). Their work highlighted the importance of multi-scale feature analysis for medical image diagnostics but was not specifically tailored to malaria detection, leaving an opportunity to adapt such techniques to the challenges of malaria diagnostics.

In this context, our research bridges these gaps by integrating object detection and instance segmentation using Mask R-CNN. By building on the strengths of Faster R-CNN and incorporating segmentation masks, our approach enables precise cell detection and subtype classification. This methodology addresses the limitations of prior work, particularly in capturing morphological details and supporting multi-class classification, paving the way for improved diagnostic precision and scalability in malaria detection systems.

3. Proposed Method

3.1. Dataset

The dataset used in this project is the BBBC041 Malaria Dataset, available from the Broad Bioimage Benchmark Collection (BBBC). This dataset contains microscopic images of red blood cells infected by the *Plasmodium falciparum* parasite. There are 6 cells in total - red blood cells and leukocytes as uninfected cells and trophozoites, schizont, ring and gametocytes as infected cells, each representing the stage of disease progression. They are annotated with bounding boxes that delineate infected and uninfected cells, enabling both object detection and classification tasks. The dataset is widely used for benchmarking computational models aimed at aiding malaria diagnosis.

3.2. Baseline Model: Faster R-CNN

As an initial baseline, we implemented the Faster R-CNN model for object detection (Hung et al., 2017). Faster R-CNN is a region-based object detector that consists of:

- Backbone CNN (typically ResNet) for feature extraction.
- Region Proposal Network (RPN) to generate candidate regions.
- ROI Pooling to extract region-specific features.
- Bounding Box Regression and Classification Heads to predict object categories and bounding boxes.

While the Faster R-CNN performed reasonably well, with high precision and recall for detecting infected and uninfected cells, it did not provide fine-grained instance-level segmentation. Furthermore, the model struggled with overlapping cells and morphological subtleties, which are critical for precise malaria diagnosis.

3.3. Proposed Solution: Mask R-CNN

To address these limitations, we implemented Mask R-CNN, an advanced extension of Faster R-CNN. Mask R-CNN enhances Faster R-CNN by adding a branch for predicting object masks in parallel with the existing bounding box regression and classification heads. This architecture enables instance segmentation, which provides pixel-level classification and better morphological analysis of cells.

3.3.1. Model Architecture

Backbone: Feature Extraction We used a ResNet-50 backbone with a Feature Pyramid Network (FPN) for hierarchical feature extraction. The FPN aggregates multi-scale features, enabling robust detection of objects at varying sizes (e.g., red blood cells of differing dimensions).

Region Proposal Network (RPN): The RPN generates proposals for regions of interest (ROIs). Each proposal is scored for objectness and refined for further processing.

ROI Align: A key novelty in Mask R-CNN is the introduction of ROI Align, which replaces the ROI Pooling operation in Faster R-CNN. ROI Align ensures sub-pixel alignment of ROIs, significantly improving mask precision.

Segmentation Branch: A fully convolutional network (FCN) branch is added to predict a binary mask for each detected object. This branch operates at the pixel level, enabling instance segmentation.

Mask R-CNN provides precise instance segmentation, offering pixel-level insights into parasite shapes critical for malaria diagnostics. The ROI Align mechanism improves feature alignment, enhancing the quality of masks and bounding boxes. Its ability to handle overlapping cells addresses a key limitation of previous methods, while the pixel-level segmentation enables detailed morphological analysis for differentiating parasite stages. By combining bounding box, classification, and mask predictions in a unified framework, Mask R-CNN delivers state-of-the-art performance in multi-task learning.

Our implementation features advanced annotation mapping by converting dataset annotations into COCO format and incorporating segmentation masks for higher granularity. Using Detectron2, we crafted a custom training pipeline with balanced datasets and augmented rare classes to address class imbalance. The evaluation framework extends beyond traditional metrics to include IoU, mAP, and Precision-Recall Curves. Unlike prior approaches, our model predicts segmentation masks for various infection stages (e.g., ring, trophozoite, schizont, gametocyte), enabling more comprehensive malaria diagnostics.

3.4. U-Net for Data Augmentation

In this step, we implemented Image-to-Image Translation using a U-Net model. The goal was to enhance or transform the original images by applying a series of data augmentations and training the U-Net to reconstruct the original images from the augmented versions. These augmentations simulate the kind of distortions and variations encountered in practical scenarios. By learning to reconstruct the original image from its augmented version, the U-Net effectively learned to handle tasks like denoising, recovering details from modified images, and adapting to transformations.

3.4.1. Model Architecture

A symmetric encoder-decoder structure with skip connections was used. The encoder extracts hierarchical features, while the decoder reconstructs the images, ensuring high-resolution outputs. This architecture follows the original U-Net design as described by (Ronneberger et al., 2015).

4. Experiment

This section outlines the computational experiments conducted to validate the approaches taken in object detection, instance segmentation, classification, and image-to-image translation.

4.1. Object Detection with Faster R-CNN

The dataset consisted of images of infected and uninfected blood cells with bounding boxes, annotated in JSON format. The data was split into training, validation, and test sets. Faster R-CNN was implemented as a baseline to classify and locate different types of blood cells.

The model utilized a Feature Pyramid Network (FPN) for multi-scale feature extraction and a Region Proposal Network (RPN) for generating object proposals. The classification head predicted object classes, while the bounding box regression head refined the proposals. Loss functions included cross-entropy for classification and Smooth L1 loss for localization. Despite achieving decent results, Faster R-CNN struggled with small object classes, prompting a transition to Mask R-CNN for better segmentation capabilities.

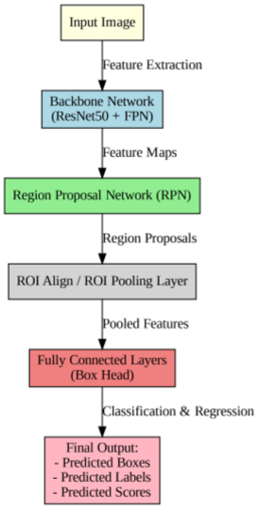

Input Image
Feature Extraction
Backbone Network (ResNet50 + FPN)
Feature Maps
Region Proposal Network (RPN)
Region Proposals
ROI Align / ROI Pooling Layer
Pooled Features
Fully Connected Layers (Box Head)
Classification & Regression
Final Output:
- Predicted Boxes
- Predicted Labels
- Predicted Scores

Figure 1. Faster R-CNN Architecture

4.2. Object Detection and Instance Segmentation with Mask R-CNN

To overcome the limitations of Faster R-CNN, Mask R-CNN was implemented for instance segmentation, adding a mask prediction branch for pixel-level segmentation. The dataset included bounding boxes and segmentation masks for six blood cell categories, following the COCO-style annotation format.

The model used ResNet-50 with FPN as the backbone, ROI Align for precise feature alignment, and multi-task loss functions for classification, bounding box regression, and mask prediction. It was trained for 1000 iterations, with evaluation metrics such as mAP, precision, recall, F1-score, and Mean IoU. Mask R-CNN demonstrated improved accuracy and interpretability, especially for smaller objects.

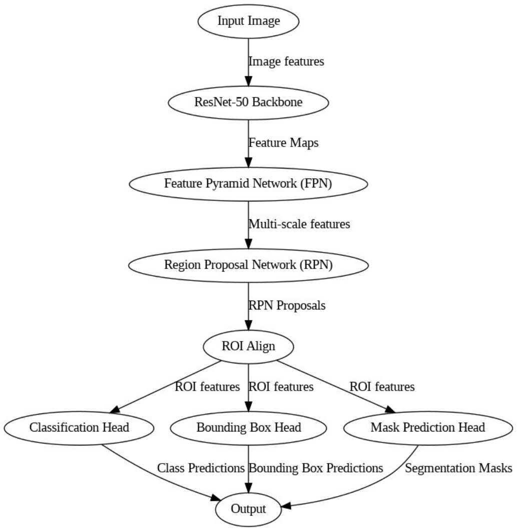


Figure 2. Mask R-CNN Architecture

4.3. Classification with MobileNet

Individual cell images were extracted based on the annotation data. MobileNet was fine-tuned to classify cropped blood cell images into six categories (mention the blood cell types from before). The dataset was augmented to address class imbalance, particularly for leukocytes. The architecture included a pre-trained MobileNet backbone, additional convolutional and dense layers, and a softmax classifier. The model achieved high accuracy with a lightweight design, making it suitable for real-time applications.

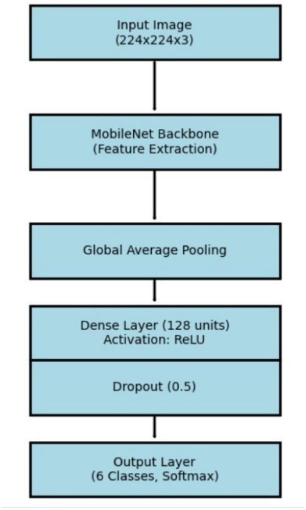


Figure 3. MobileNet Architecture

4.4. Classification with VGG-16

A pre-trained VGG-16 model was fine-tuned for multi-class classification. The architecture was adapted by adding custom top layers for the task. Training involved two stages: freezing and unfreezing layers, with data augmentation applied to improve generalization. Grad-CAM was used for explainability, providing visual insights into the model's predictions. Performance was validated using metrics like accuracy, confusion matrix, and training curves.

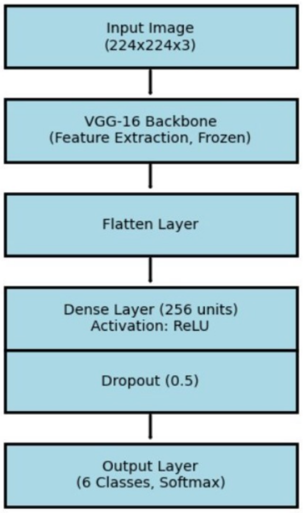


Figure 4. VGG-16 Architecture

4.5. Image-to-Image Translation with U-Net

For image-to-image translation, U-Net was trained to reconstruct original images from augmented versions. The dataset included various transformations to simulate real-world variations. The U-Net architecture employed an encoder-decoder structure with skip connections, preserving both global context and fine details.

The model was trained for 50 epochs using Mean Squared Error (MSE) loss. The translated images were subsequently used as inputs for the Mask R-CNN model, ensuring its robustness across varied image distributions. U-Net's ability to generalize to diverse augmentations demonstrated its scalability and effectiveness for image translation tasks.

These experiments validate the methodologies adopted for blood cell analysis, addressing object detection, segmentation, and classification challenges in medical imaging.

5. Results

The experiments yielded insights into the performance of MobileNet, VGG-16, and Faster R-CNN models for malaria cell analysis, including classification and object detection tasks.

5.1. Faster R-CNN Results

The Faster R-CNN model was evaluated on object detection tasks with the following metrics: - mAP (Mean Average Precision): 0.3824 - mAP (50% IoU): 0.5072 - mAP (75% IoU): 0.4657

Interpretation of Results:

- Loss Metrics: Training loss consistently decreased, indicating effective learning from the training data. However, validation loss fluctuated, suggesting potential overfitting and limited generalization to unseen data.

- mAP Metrics: The overall precision-recall balance across all IoU thresholds was low, reflecting challenges in detecting objects with high precision. The results highlight difficulty in achieving accurate bounding boxes, particularly for small objects.

Possible Reasons for Performance Limitations:

1. Small Dataset Size: The dataset of 1328 images was insufficient for a complex object detection task, especially with diverse classes and annotations.

2. Class Imbalance: Dominance of certain classes led to suboptimal performance for minority classes.

3. Small Object Detection Challenges: Faster R-CNN struggled with detecting small objects due to limited pixel information.

4. Annotation Quality: Inconsistent bounding box annotations may have contributed to reduced performance.

5. Model Configuration: The ResNet50 backbone and anchor box configurations were potentially suboptimal for the dataset.

Suggestions for Improvement:

- Increase dataset size through collection or augmentation.

- Address class imbalance using oversampling, weighted loss, or synthetic data.

- Enhance small object detection with deeper backbones (e.g., ResNet152) or tailored anchor box configurations.

- Improve annotation quality and diversify training data through robust augmentation strategies.

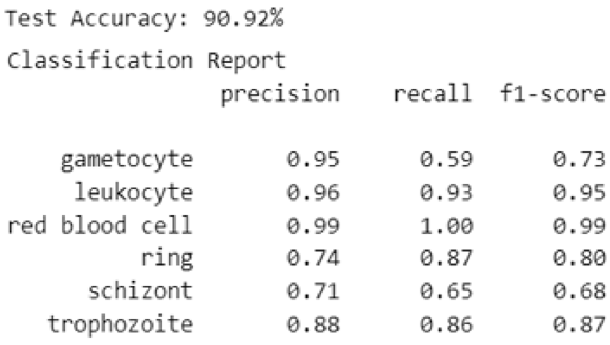
- Fine-tune hyperparameters like learning rate, batch size, and loss functions.

Future Potential:

With these enhancements, Faster R-CNN could achieve mAP values in the range of 0.6–0.75, with mAP (50% IoU) values exceeding 0.8. This would make it more effective for small-object detection in malaria cell imaging. Hence, as discussed before, we plan to solve these shortcomings using Mask R-CNN which adds a mask prediction head thus also performing instance segmentation providing pixel-level masks for each detected object.

5.2. MobileNet Results

The MobileNet model achieved a classification accuracy of 90.92%, demonstrating its effectiveness in classifying malaria subtypes, excelling the results showed in (Cheuque et al., 2022). Its lightweight architecture and efficient feature extraction made it particularly suitable for real-time diagnostic applications. Data augmentation techniques, such as rotation, flipping, and brightness adjustments, helped mitigate class imbalance and enhance model robustness. These results highlight MobileNet as a scalable and practical choice for malaria diagnosis tasks.



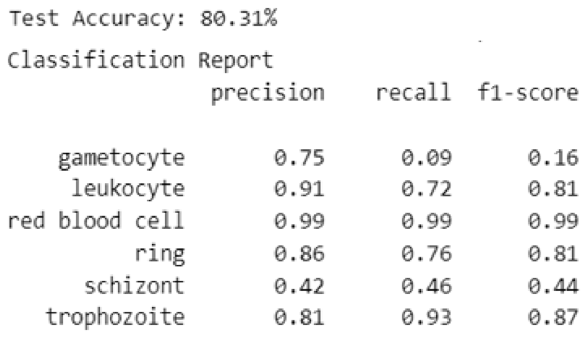
```
Test Accuracy: 90.92%
Classification Report
                precision    recall  f1-score

    gametocyte       0.95      0.59      0.73
     leukocyte       0.96      0.93      0.95
red blood cell       0.99      1.00      0.99
          ring       0.74      0.87      0.80
      schizont       0.71      0.65      0.68
   trophozoite       0.88      0.86      0.87
```

Figure 5. MobileNet Model Metrics

5.3. VGG-16 Results

The VGG-16-based model provided stable performance, utilizing a pre-trained backbone for feature extraction. While slightly less efficient than MobileNet, the model achieved reliable training and validation outcomes. The integration of Grad-CAM added interpretability, allowing for visual explanations of the model's predictions. Despite its heavier architecture, VGG-16 demonstrated strong generalization for multiclass classification, making it a valuable option for tasks prioritizing transparency over computational efficiency.



```
Test Accuracy: 80.31%
Classification Report
                precision    recall  f1-score

   gametocyte        0.75      0.09      0.16
    leukocyte        0.91      0.72      0.81
red blood cell       0.99      0.99      0.99
         ring        0.86      0.76      0.81
     schizont        0.42      0.46      0.44
  trophozoite        0.81      0.93      0.87
```

Figure 6. VGG-16 Model Metrics

5.4. Comparison and Implications

MobileNet outperformed VGG-16 in terms of efficiency and accuracy, establishing it as the preferred model for real-time classification. Meanwhile, Faster R-CNN demonstrated promise in object detection but faced challenges due to dataset limitations and model constraints. These findings underscore the need for optimized datasets, advanced architectures, and robust preprocessing techniques to maximize the effectiveness of deep learning in medical imaging.

These results lay the groundwork for future advancements, such as integrating segmentation tasks or exploring ensemble methods to improve accuracy and reliability in malaria diagnosis.

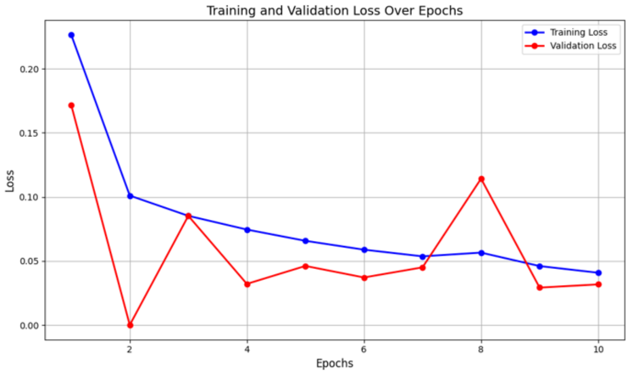


Figure 7. Training and Validation Loss over Epochs for the Faster R-CNN Model

5.5. U-Net Results

The U-Net model was trained for 10 epochs to perform image-to-image translation, reconstructing original images from their augmented counterparts. While the model demonstrated the ability to learn the general mapping between augmented and original images, the results are not yet optimal, as evidenced by the reconstructed images (examples shown in Figure 8). Training is still in progress, and additional epochs are required for the model to converge and achieve higher-quality outputs.

The current results show that the model captures some structural details but struggles with finer features, likely due to the early stage of training. The use of a robust augmentation pipeline, including random flips, rotations, and color jitter, is expected to help the model generalize better to diverse input variations as training progresses.

The reconstructed images, despite being incomplete, show promise for further refinement and will serve as an essential input for the Mask R-CNN segmentation task. Future training iterations, with extended epochs and potential fine-tuning, are expected to yield higher-quality results, enabling more accurate downstream analysis.

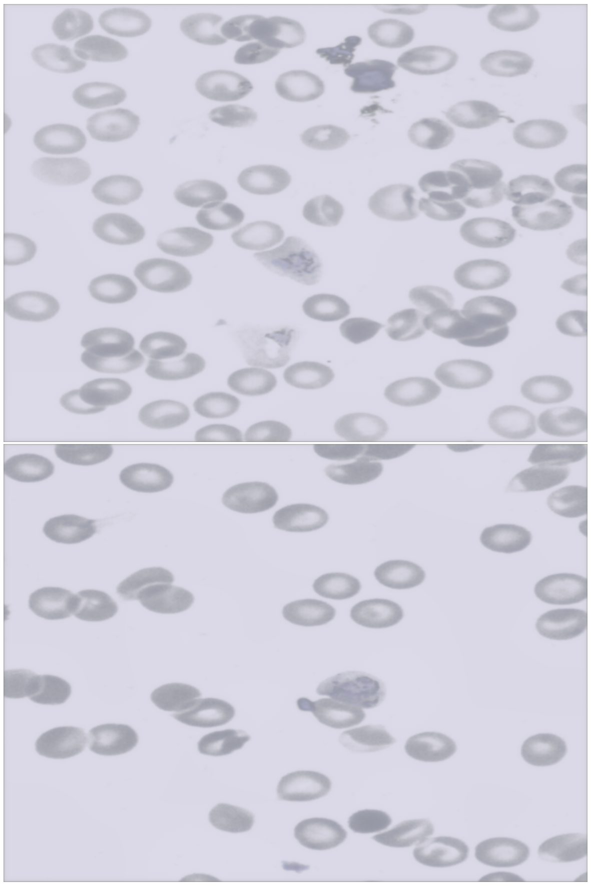


Figure 8. Example of reconstructed images generated by the U-Net model after training on augmented data.

6. Discussion

This study highlights the potential of deep learning models like MobileNet, VGG-16, and Faster R-CNN for malaria diagnosis through classification and object detection tasks.

However, limitations such as small dataset size, class imbalance, and challenges in small object detection impacted performance for Faster R-CNN and U-Net. Improving data quality and diversity, leveraging advanced architectures, and fine-tuning hyperparameters could address these issues. Mask R-CNN Model is sill running on Turing which is why we were unable to obtain the results. Images translated by U-Net could have been of higher quality but they were trained only on 10 epochs due to extremely long training time.

The results suggest that lightweight architectures like MobileNet are well-suited for real-time diagnostic applications, while more complex models like Mask R-CNN offer promise for segmentation tasks.

7. Conclusions and Future Work

This study demonstrates the potential of deep learning models for malaria diagnosis through classification and object detection tasks. MobileNet achieved the highest accuracy (90.92%) for multiclass classification, highlighting its suitability for real-time applications. VGG-16 provided robust performance and interpretability, while Faster R-CNN effectively localized objects but struggled with small objects and dataset limitations.

For future work, the Mask R-CNN model will be trained using data augmented with the U-Net model, enabling robust instance segmentation and improving performance on modified image distributions. This approach aims to address challenges in object detection and enhance segmentation accuracy. Additionally, further exploration of ensemble models, advanced data augmentation techniques, and deployment in clinical settings will pave the way for more effective and scalable diagnostic tools.

References

Cheuque, Claudio, Querales, Miguel, León, Ricardo, Salas, Ricardo, and Torres, Roberto. An efficient multi-level convolutional neural network approach for white blood cells classification. *Diagnostics*, 12(2):248, 2022. doi: 10.3390/diagnostics12020248.

Chima, Jaspreet Singh, Shah, Abhishek, Shah, Karan, and Ramesh, Rekha. Malaria cell image classification using deep learning. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(6):5553–5556, 2020. doi: 10.35940/ijrte.F9540.038620.

Hung, Jane, Lopes, Stefanie C. P., Nery, Odailton Amaral, Nosten, Francois, Ferreira, Marcelo U., Duraisingh, Manoj T., Marti, Matthias, Ravel, Deepali, Rangel, Gabriel, Malleret, Benoit, Lacerda, Marcus V. G., Renia, Laurent, Costa, Fabio T. M., and Carpenter, Anne E. Applying faster r-cnn for object detection on malaria images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 808–813. IEEE, 2017. doi: 10.1109/CVPRW.2017.112.

Ronneberger, Olaf, Fischer, Philipp, and Brox, Thomas. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28. URL https://arxiv.org/abs/1505.04597.