

Deep Learning Based Automatic Identification of Malaria Infected Cells

Pothuraju V. V. Satyanarayana¹, Kanakala Sai Kumar^{2*}

¹Associate Professor, Department of Computer Applications, Visakha Institute of Engineering & Technology(A), Visakhapatnam, India

²PG Research Scholar, Department of Computer Applications, Visakha Institute of Engineering & Technology(A), Visakhapatnam, India

Abstract: The study focuses on the application of artificial intelligence as a way of classifying malaria-infected red blood cells accurately and transparently. In particular, the classic method of diagnosing malaria is centered on a painstaking examination of the blood smear by qualified personnel which still proves effective, but time- and labor-consuming, as well as subjective, and dependent on the presence of professionals. Nevertheless, deep learning provides an information-driven approach that allows for the identification of unique visual patterns in labeled cells. Thus, the aim of this study is to develop and assess various models based on convolutional neural networks that would help distinguish the categories of red blood cells based on blood smear images into Parasitized and Uninfected ones. The methodology implies collection of the data set with images of red blood cells, their processing, and transfer learning on the basis of models like ResNet and DenseNet. Additionally, the use of the Grad-CAM technology helps make the diagnostic system even clearer. The findings indicate that DenseNet121 greatly exceeds the performance of ResNet50 with a test accuracy of 92.15% as opposed to the 70.19% accuracy of ResNet50, thereby showing the benefits of densely connected feature re-use in detecting minute intracellular parasitic features.

Keywords: Malaria Detection, Deep Learning, Convolutional Neural Networks, Grad-CAM, Explainable AI, Transfer Learning, Medical Image Classification.

1. Introduction

Diagnosis of malaria [1] has always been regarded as significant due to its great impact on human life especially in tropical and sub-tropical regions where malaria continues to be a big public health issue [2]. It is crucial to be able to determine the presence of Plasmodium-infected red blood cells in order to treat the disease. Historically, malaria diagnosis depended on manual microscopic analysis of Giemsa-stained thin blood smears, which, although the clinical standard for diagnosis, is laborious, needs special skills, and suffers from variability between examiners. Accordingly, that has created the need for computer-supported approaches that would complement the existing diagnostic methods [3].

In recent times, artificial intelligence (AI) and deep learning were introduced into biomedical image analysis. In contrast to traditional image processing guided by rules, deep learning methods are data-oriented by use of labelled clinical images [4]. Therefore, convolutional neural networks are suitable for

malaria cell recognition due to their ability to process complex visual information.

Integrating deep learning into malaria diagnostics not only increases screening efficiency but also provides important societal benefits. Automated classification can enable healthcare facilities in high-burden, low-resource areas to effectively evaluate high patient volume and shorter diagnosis time as well as facilitate pathologists in their work by flagging cases for follow-up [8]. Additionally, the use of explainability methods such as Grad-CAM [11] makes the model more trustworthy for practitioners due to visualizing the process of arriving at a particular decision.

Although the choice of model to use is critical, the issue of training data that is both sufficient and diverse is also of utmost importance in making a diagnosis [14]. Domain shifts may arise as a result of changes in staining density, optical microscope resolution, illumination conditions, and methods of slide preparation in different laboratories. Thus, the generalization of the model will be compromised unless measures are taken to eliminate these potential concerns. To do away with these issues, various types of data alteration are employed in practice, including rotation, flipping, or changing illumination conditions [3]. Other than that, it is necessary to preprocess data properly in order to make sure that important parasite indicators receive due attention and unimportant classifiers do not obscure such indicators. Moreover, another thing that should be done is to conduct modeling on CPU-equipped machines in situations when technical capabilities are limited.

In this study, the authors present a methodology for building deep learning systems aimed at identification of malaria cells from images and classification them to two categories, i.e. Parasitized and Uninfected [14]. Using transfer learning, CPU-friendly training, and Grad-CAM-based explanation, the researchers strive to create a comprehensive and comprehensible classification pipeline. The results show that DenseNet121 is more capable of identifying subtle parasitic signs than ResNet50, thus providing an opportunity for the development of intelligent and efficient malaria diagnostics systems.

*Corresponding author: kanakalasaikumar00@gmail.com

2. Literature Review

Das et al. (2013) [3] proposed a machine learning technique to automate the identification of malaria parasites through light microscopic image analysis, using 19 features related to shape, size and texture of erythrocytes. The study used f-statistic based feature selection and combined it with Bayesian learning and support vector machine algorithm. The results of the study showed that Bayesian technique resulted in the largest hit rate, which was 84% using 19 features, while the SVM algorithm achieved only 83.5% accuracy by using 9 features.

Similarly, Liang et al. (2016) [4] studied the role of convolutional neural networks in the diagnosis of malaria. The purpose of the study was to develop a customized CNN that can be used to diagnose malaria. The authors used a custom-built neural network with 16 layers that have been trained end to end on the same dataset. The authors performed 10 fold cross validation for the dataset of 27,578 single-cell images, and the findings demonstrated that the model developed using CNN was very accurate, reaching the hit rate of 91.37% and beating a transfer-learning approach that achieved only 91.29% accuracy using the same dataset.

The work done by Rajaraman and others (2018) involves the evaluation of different architectures of pre-trained CNNs such as AlexNet, VGG-16, ResNet-50, Xception, and DenseNet-121 in extracting features for malaria diagnosis. They also evaluated the models under fully fine-tuned modes and using GPU, with ResNet-50 achieving an accuracy of up to 0.959, with DenseNet-121 achieving an accuracy of 0.952, with the highest sensitivity of 0.960 compared to the models they have tested. The less significant gap between the performances of the two architectures under the fully-trained situation compared to the broader gap observed in the present research indicates that the underperformance of ResNet50 is closely linked to the limitation of the training budget rather than its architectural features.

In their research, Bibin et al. (2017) introduced a deep knowledge network comprising six layers and built by stacking restricted Boltzmann machines and combining features of color and texture. They reported the achieved F1 score of 89.66%, sensitivity equals to 97.60%, and specificity of 95.92% on a small dataset of 4100 images. Despite the relative speed of achieving these results, they have not outperformed DenseNet121 F1 score (equal to 0.9241) achieved in this paper, which emphasizes the importance of CNN-based architectures using transfer learning compared to previous techniques.

Both studies by Fuhad et al. (2020) [13] and Sriporn et al. (2020) [12] state that the accuracy achieved was greater than 98% (99.23% and 98.86%, respectively). Their methods permitted such successful results through using techniques specialized in achieving optimal functioning – knowledge distillation, auto-encoder training, and tuning activation functions (Mish, Nadam) along with full multi-epoch training. This report does not consider these achievements to be entirely comparable but rather sees the huge difference between the research of the authors and its one-epoch, CPU-based results to be clear evidence of the compromise achieved between full optimization of the methods and actually possible practical

implementation of the methods developed in low-resource clinical practices.

The studies mention that it is possible to achieve an accuracy of 95–99% when using a fully optimized, GPU-based CNN approach, while the result achieved might depend on a huge computational budget that is hardly available in the clinics suffering from malaria the most. This paper aims to compare ResNet50 and DenseNet121 in one more straightforward way – in the one-epoch, CPU experiment and hopes to prove that DenseNet121 has highly-developed feature reusing that may ensure important benefits from using the method.

As a further extension of comparative analysis in literature, in their 2025 research, Yakhshiboev et al. [15] studied eight different pretrained CNN models including Xception, ResNet-50, and ResNet-101 to classify malaria parasites based on microscopic blood smear imagery. It was found that the study resulted in achieving the accuracy of not more than 89%, while the authors emphasize that the use of transfer learning has great implications for improving accuracy thus considerably reducing time required for training in comparison to the earlier diagnosis methods. The results demonstrate that the benefits of transfer learning depend on the specific architecture of the used CNN model and its use is not guaranteed simply because the pretrained model is used.

One more similar finding was obtained by Pramudito et al. (2022) [16], who analyzed ResNet50V2, InceptionV3, and EfficientNetB0 based on red blood cell smear images without modifications in model's architecture. According to their research, ResNet50V2 showed an accuracy of 0.76–0.81, EfficientNetB0 demonstrated 0.76–0.80 and InceptionV3 achieved 0.77–0.82. The results expose the importance of applying proper finetuning since all the known variants based on ResNet features vary in their overall performance in relation to the dataset studied, model preprocessing issues, and finetuning approaches.

Diving deeper into ensemble and hybrid techniques, Alqahtani et al. (2025) [17] presented an ensemble which is a combination of a tailored CNN with VGG16, ResNet50V2, DenseNet201, and VGG19. The ensemble reached a testing accuracy of 97.93%, which is greater than that of the separated models; feature-extraction techniques using CNN (97.20%) and VGG16 (97.65%) as well as the hybrid model CNN-SVM (82.47%). The difference shows that the hybrid feature-extraction methods do not automatically surpass their end-to-end fine-tuned deep architectures and ensemble methods of applying the constitution of several fully trained backbones have a completely different computational possible than using a single model with limited resources.

The landmark study by Rajaraman et al. was conducted on the NIH malaria cell dataset [18] and proved that transfer learning from the natural images collection significantly exceeds results obtained from scratch training. Later experiments based on this study, for instance, hybrid 12-layer CNN together with EfficientNet-B7 recorded accuracy level of 97.96% working with a more extensive dataset of 43400 images and highlighting that surpassing results in accuracy are explained not only by the nature of the chosen architecture.

In conclusion, older methods based on feature extraction provide important perspectives on current CNN designs. Feature extraction techniques using deep belief networks and tailored features, as examined by Das *et al.* (2013) and Bibin *et al.* (2017), provided varying results when measured with CNN extraction methods, which depended on the backbone architecture, dataset size, and complexity of training in relation to the overall success of the model during malaria parasite classification.

A. Research Gap

Despite the substantial contributions deep learning has made to automating the diagnosis of malaria through blood smear image analysis, most studies providing deep learning for malaria diagnosis have focused on achieving top performance levels in artificial conditions that are impossible to replicate in real-world clinics where most cases occur. While such studies are methodologically interesting, the results they produce do not provide insights into how such architectures would perform in low-resource clinics and field laboratories where GPU devices are not available, which is commonly the case. In fact, most studies do not answer the questions about the performance of deep learning models after being trained under budget restrictions of one epoch, despite the fact that this budget is indeed a reflection of the reality where such tools might be applied. Unfortunately, most of the figures indicated in the literature are validated figures but not true held-out figures, though this important distinction is often forgotten. With validation datasets often used for hyperparameter tuning, the validation accuracy may exaggerate the model's true transferability and make cross-study comparisons less reliable than they seem.

Another gap exists around the comparative performance of various CNN architectures in similar settings. Much of the existing literature is oriented towards analysing optimization type rather than architectural nature, emphasizing that numerous approaches including multiple combinations of pretrained or mixed CNN structures have been merged in order to achieve higher accuracy. While this strategy may indeed work well, it conceals a much more important question of how different structures such as ResNet and DenseNet differ in their capability to extract diagnostically significant features given limited training time and resources. The comparative research that has been done involves testing a variety of architectures under full training conditions, when the differences in structures become less relevant, and report performance in validation set as though this represented the ultimate indicator of their diagnostic reliability. As a result of this approach, evidence on the factors influencing performance of different architectures such as use of retained features in DenseNet or skip connections in ResNet is missing in low-resource and low-epoch training scenario research involving observations with non-existing initial features.

In conclusion, the issue of interpretability in malaria diagnosis models remains under-investigated compared to their predictive capacities. Though visualizing model behavior in the form of tools such as Grad-CAM has been suggested, predictive

performance is often prioritized over the transparency of diagnosis models in most high accuracy studies. This is a particularly relevant concern in medical diagnostic applications, where physicians are unlikely to employ models lacking any visual or other type of explanation, even where verification accuracy is high. Given these problems, it becomes clear there is a pressing need for a cost-efficient, one-epoch classification system that serves the purpose of comparing, on the same conditions, ResNet50 and DenseNet121 models using a separate test dataset along with the application of Grad-CAM method of interpretability, thus allowing the acquisition of a complete and clinically applicable understanding of the functioning of transfer learning methods in the context of real-life challenges of implementing malaria diagnostics.

3. Objective of Work

A. Problem Statement

Malaria detection relies on well-known methods such as human microscopy of blood smears stained with Giemsa solution. Despite the clinical validation of this method, it is known to be slow, prone to differences between practitioners, and reliant on experts. The fact that this last group is scarce in many tropical and subtropical regions means that innovations based on machine learning have a lot of potential for automatization of the process. Nevertheless, most current classifiers work on the principles of CNN, meaning that their usage for practical purposes is limited due to high hardware requirements, making their use for the most under-resourced hospitals impossible. Moreover, most of the models currently available do not explain their decisions in a clear way too, making it hard for them to become appealing for practitioners. The problem is therefore formulated as the need to develop and evaluate a transfer learning-based classification framework, trainable under realistic CPU-only, single-epoch constraints, that can accurately distinguish Parasitized from Uninfected red blood cells while also providing visual explainability of its decision-making process through techniques such as Grad-CAM.

B. Research Objective

To develop, train, and assess deep-learning-based models for the automated grouping of malaria-infected red blood cells into Infected and Uninfected groups, using microscopic blood smear images is the main aim of this research. The study will preprocess the malaria cell image dataset to standardize and normalize its image size, and quality, leverage the transfer learning approach based on pretrained convolutional neural networks in terms of ResNet50 and DenseNet121 architectures, and evaluate their performances comparatively under realistic conditions of running on only one epoch of one CPU. Further, the model performance will be assessed using the results of the test set, rather than solely relying on the results of the validation set, analyzing such metrics as accuracy, precision, recall, and F1 score. Moreover, the incorporation of the Grad-CAM methods of visualization will also be a part of the research, in order to make the model predictions and classifications

interpretable and understandable in the clinical context.

4. Methodology and Architecture

In this research work, a well-defined strategy of transfer-learning has been utilized to classify the images of blood cells into two categories: parasitized and non-infected, with the help of microscopic blood smear data. The first step of this process involves collecting data from the labeled dataset of red blood cells, where the images were separated into different datasets for training and testing to mitigate the chances of data bleeding over from one phase to the other. The next step involves preprocessing of the images, in which the images are converted into fixed dimensions of 96×96 pixels. Pixel densities need to be standardized as well, and hence, for normalization purposes, the pixel densities are made to fall in between the range of $[0,1]$ by adjusting the raw pixel values of their corresponding images.

Once the preprocessing is done, the ResNet50 and DenseNet121 convolutional neural network architectures are implemented. Both models benefit from the technique known as transfer learning, as their layers are initialized based on their training on previously established dataset(s) that consist of various images taken from nature (and this way, they will not have to learn the basic properties of the object). The two structures used are tested under identical experimental conditions, which means using only CPU and processing the data for a single epoch, and thus the results are based on the characteristics of the two models but not on any advantages related to hardware facilities. This comparison is an important part of this writing because it aims to answer the question about the difference between ResNet50 approach and DenseNet121 approach when training under the computer restrictions.

A separate test set not seen by the model while training or tuning hyperparameters. Some of the evaluation metrics are overall accuracy, precision, recall, and F1-score calculated separately for both classes namely Parasitized class and Uninfected class to check any performance imbalance related to the class. Apart from these evaluation metrics, a number of visual diagnostic measures will be created to get a broader outlook on how the model is behaving, such as the normalized confusion matrices which indicates the number of true positives, true negatives, false positives, and false negatives; Receiver Operating Characteristic (ROC) curves along with the Area Under Curve (AUC) value to evaluate the discriminative nature of the model in terms of the classification thresholds; and precision-recall curves which matter significantly due to the importance of reducing false negatives in the diagnosis of diseases.

The entire system architecture is divided into four distinct modules: Data Layer responsible for images' preprocessing; Model Layer with the two independent CNNs; Evaluation Layer which calculates metrics and produces Grad-CAM visuals; and Presentation Layer created as a simple GUI that allows playing with single images, comparing results of models' operation, and generating pre-determined charts and visuals. The logical separation of the aspects provides reproducibility, makes debugging process easier, and gives a possibility to add new models and evaluation methods in the

future without needing to change a lot in the working code.

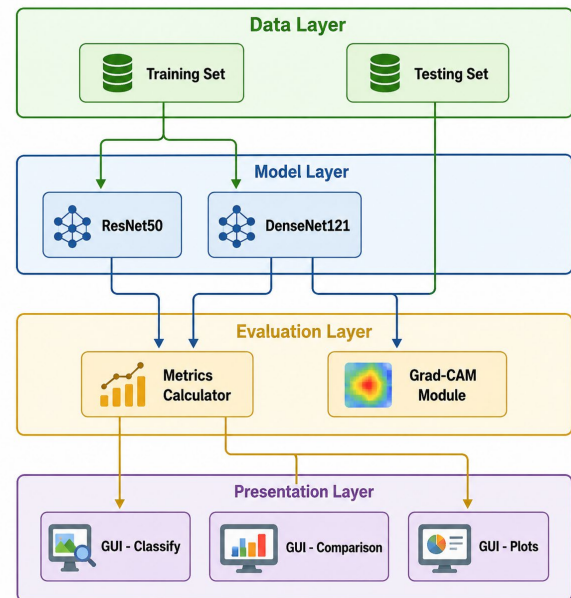


Fig. 1. System architecture diagram

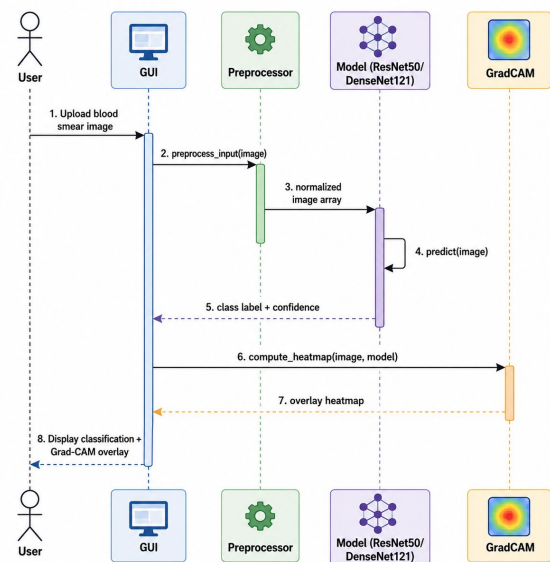


Fig. 2. Classification workflow sequence diagram

First, a participant uploads an image of a blood smear through the GUI. This image is fed into the preprocessing module, responsible for resizing it to 96×96 pixels and standardizing pixel values to a range of $[0, 1]$. The next step involves feeding the processed image to the chosen model, ResNet50 or DenseNet121, which infers upon the image and comes up with a predicted class (parasitized vs uninfected) alongside a confidence level. Next, the image and model are passed to Grad-CAM module which calculates gradients from the last layer used for convolution in order to create a heat map showing the most significant regions that led to the prediction. This heat map is then superimposed upon the initial image. In the end, the GUI shows the classification result, confidence level and the Grad-CAM heat map to the participant for

effective diagnosis.

5. Findings and Results

Analysis of the results of the comparison of ResNet50 and DenseNet121 taught under the same CPU-centric single-epoch settings showed that there was a serious performance difference between the two systems. DenseNet121 demonstrated a testing accuracy value of 92.15%, as opposed to the result of 70.19% achieved by ResNet50. This difference proves that even in the framework of the experiment with the same conditions for training, DenseNet121, due to the dense connection of architecture, allows for better utilization of the features learned, while ResNet50, which is based on the idea of residual connections, could potentially reach such performance only after many epochs of training.

The findings from the evaluation of DenseNet121 and ResNet50 were corroborated by the metrics of precision, recall, and F1 score. DenseNet121 had an impressive F1 score of 0.9241 indicating that it was able to maintain both high sensitivity and specificity without sacrificing one for the benefit of the other. In contrast, ResNet50 demonstrated poorer performance showing poorer classification results across one of the two classes. The normalized confusion matrices provided evidence for the conclusion of the metrics evaluation of the two models as well: while there was a strong concentration of predictions along the diagonal in the case of confusion matrix of DenseNet121 meaning that most of the test instances were predicted correctly, the confusion matrix for ResNet50 had a more diverse pattern resulting in higher number of false positives and false negatives.

The AUC values and ROC curves demonstrated the presence of an existing performance gap. DenseNet121 achieved relatively good AUC values, demonstrating a high performance level with a significant ability to distinguish between labels, unlike ResNet50, whose curve seemed to be positioned close to the diagonal line of reference. Similar trends could be seen in precision-recall curves while DenseNet121 demonstrated very high precision under a varied recall circumstance, ResNet50 showed a poor capacity to separate between infected and healthy classes when it comes to practical implementation of the network as a medical diagnostic tool.

In turn, based on confidence distribution plots, it can be observed that DenseNet121 yields predictions with a lot of confidence, with most of the predicted probabilities gathered quite close to the 0 and 1 values of the distribution. As for ResNet50, its predictions distribution can be seen concentrated around the decision boundary. The overall accuracy assessment for both classes has confirmed the previous findings according to which DenseNet121 obtained pretty good performance level along with ResNet50 showing a considerable uneven distribution of the performance level for one class at the cost of another class.

In the final analysis, Grad-CAM visualizations produced interesting qualitative impressions related to model reasoning process. DenseNet121 usually produces heat maps localized in parts of the cell corresponding to visible forms of parasites. Hence, it can be interpreted that the results produced by the

model correspond to clinically important morphological features of the cell studied. decision-making models that matter for clinical practice.

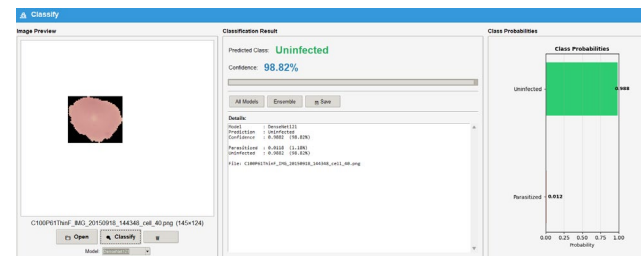


Fig. 3. Malaria cell classification Using DenseNet121

This screen shows the classification result of a blood cell image using the DenseNet121 model. The uploaded image is displayed on the left, while the center panel presents the predicted class as Uninfected with a confidence score of 98.82%.

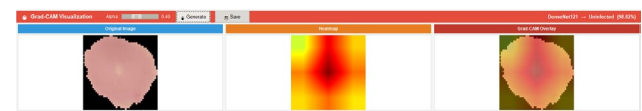


Fig. 4. Grad-CAM visualization of DenseNet121 prediction

This screen presents the Grad-CAM visualization for the DenseNet121 model's prediction. The left panel displays the original blood cell image, the middle panel shows the generated heatmap highlighting regions that influenced the model's decision, and the right panel overlays the heatmap on the original image for better interpretation. The model predicts the cell as Uninfected with 98.82% confidence, providing visual evidence to improve the transparency and interpretability of the classification process.

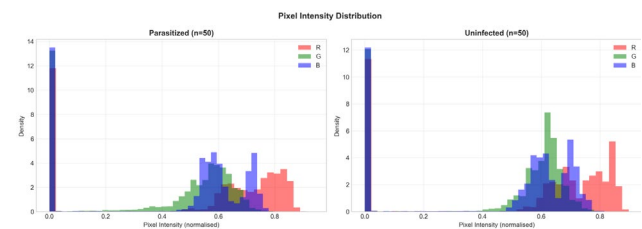


Fig. 5. Pixel intensity distribution of malaria cell images

The figure shows the normalized RGB pixel intensity distributions for 50 parasitized and 50 uninfected malaria cell images. Both classes exhibit a prominent peak near zero, representing the black background surrounding the segmented cells. The parasitized images display a wider spread of red, green, and blue intensities, indicating greater color variation caused by parasite presence and Giemsa staining. In contrast, uninfected cells have more compact intensity distributions, reflecting their uniform appearance. Although the distributions overlap, these color differences provide useful low-level features that convolutional neural networks exploit for accurate malaria cell classification.

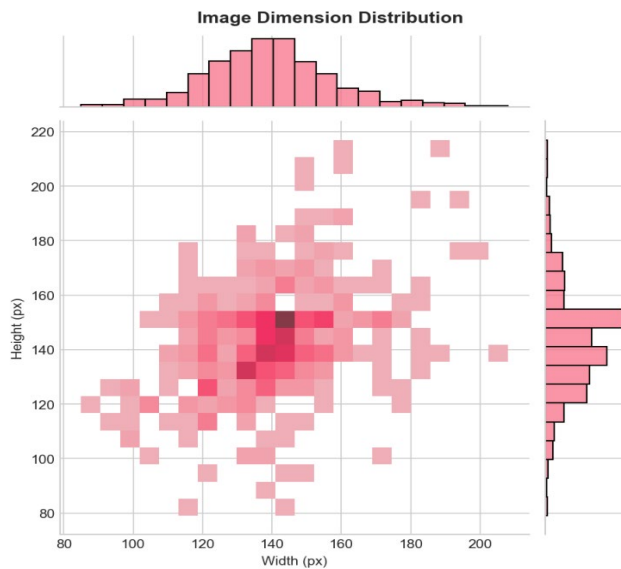


Fig. 6. Distribution of image width and height in the malaria cell dataset

Most images are concentrated around 140×145 pixels, indicating a consistent image size across the dataset. The width ranges approximately from 80 to 210 pixels, while the height varies from 80 to 220 pixels, with only a few outliers. The clustered distribution suggests minimal variation in image dimensions, reducing preprocessing complexity. This consistency facilitates uniform resizing to the model input size (224×224 pixels), improving training efficiency and ensuring reliable feature extraction by the convolutional neural network.

6. Conclusions

This research aimed to fill an important gap in the literature about malaria diagnostics: the comparative analysis between different deep learning architectures in a controlled manner in conditions that are similar to those in the low-resource settings in terms of technology. The study was conducted by training ResNet50 and DenseNet121 in exactly the same way, using the same training methods for both of the models, and then the performance of both models was evaluated based on the appropriately selected test data. The findings of the study clearly indicate that the performance of DenseNet121 was considerably better than that of ResNet50, resulting in a 92.15% test accuracy of DenseNet121, while ResNet50 showed only 70.19%. Additionally, the performance of DenseNet121 measured in terms of precision, recall, F1 score, AUC, and the degree of confidence is much better than that demonstrated by ResNet50. This finding is explained by the fact that DenseNet121 is able to reuse the features from previous layers due to the fact that it is designed in the way that enables feature reuse, and thus, it is more suitable for rapid adjustment to the low-resource setting compared to ResNet50.

Apart from pure predictive performance, the employment of Grad-CAM-based explainability provides another important clinical dimension pertaining to this work. The heatmaps created by DenseNet121 indicate a greater tendency to focus on biologically relevant parasitic structures so that it can be deemed interpretable enough to promote clinical trust, which is

often neglected in malaria diagnostic studies of high-performing devices that are more focused on matching benchmarks than providing transparency.

All these findings suggest that DenseNet121 is more suitable for being employed in malaria diagnostic devices operating in low-resource CPU-only environments where training duration and computational capacity are limited.

7. Limitation

The principal limitation of this research lies in the one-epoch training limitation imposed on the outcomes of the research as a result of which neither of the models is able to show their highest possible accuracy hence rendering the comparisons with trained models from literature unjustified. Though the data used is balanced, its limited data sources and staining protocols leads to the decrease of its generalizability as applicable to blood smears obtained in other laboratories or using different imaging methods. Furthermore, the study deals with a binary classification only neglecting the species identification of the parasite and infection stage classification which are of considerable importance in making clinical decisions. The last limitation involves the lack of validity of the qualitative assessment according to the Grad-CAM.

8. Future Work

Next research could broaden the basis of comparison to include various other structures; including lighter models such as MobileNet or EfficientNet; in order to discover additional possibilities that will work well when resources are limited. Increasing number of training epochs continuously can assist in determining the point where differences in architecture are diminished and the price of training is justified so as to achieve accuracy. Expanding the database and including various clinical sources and staining protocols will make the tool more generalizable, while advancing from binary classification to classification of different species of parasites and infection stage will make it more clinically relevant. Finally, confirming the Grad-CAM results based on ground truth provided by specialists can make the results more interpretable and support real-life adoption.

References

- [1] World Health Organization, *World Malaria Report 2025*, Geneva, Switzerland: World Health Organization, 2025.
- [2] World Health Organization, "Malaria fact sheet," *WHO Fact Sheets*, Nov. 2025.
- [3] D. K. Das *et al.*, "Machine learning approach for automated screening of malaria parasite using light microscopic images," *Micron*, vol. 45, pp. 97–106, Feb. 2013.
- [4] Z. Liang *et al.*, "CNN-based image analysis for malaria diagnosis," in *Proc. IEEE Int. Conf. Bioinform. Biomed. (BIBM)*, 2016, pp. 493–496.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
- [6] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4700–4708.
- [7] Y. Dong *et al.*, "Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2017, pp. 101–104.

- [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, Apr. 2018.
- [9] A. Rahman *et al.*, "Improving malaria parasite detection from red blood cell using deep convolutional neural networks," in *Proc. IEEE Region 10 Symp. (TENSYP)*, 2019, pp. 234–239.
- [10] A. Vijayalakshmi and B. Rajesh Kanna, "Deep learning approach to detect malaria from microscopic images," *Multimed. Tools Appl.*, vol. 79, no. 21, pp. 15297–15317, Jun. 2020.
- [11] R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.
- [12] K. Sriporn *et al.*, "Analyzing malaria disease using effective deep learning approach," *Diagnostics*, vol. 10, no. 10, p. 744, Sep. 2020.
- [13] K. M. F. Fuhad *et al.*, "Deep learning based automatic malaria disease detection," *Information*, vol. 11, no. 5, p. 247, May 2020.
- [14] National Library of Medicine, "Malaria cell images dataset," U.S. National Institutes of Health, 2018. [Online]. Available: <https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html>
- [15] M. Yakhshiboev *et al.*, "Malaria parasite cell classification using transfer learning with state-of-the-art CNN architectures," *Biology*, vol. 14, no. 12, p. 1792, Dec. 2025.
- [16] A. Pramudito *et al.*, "CNN based transfer learning for malaria parasite detection using thin-blood smear images," in *Proc. 7th Int. Conf. Sustain. Inf. Eng. Technol. (SIET)*, 2022, pp. 112–117.
- [17] S. Alqahtani *et al.*, "An optimized transfer learning approach integrating deep convolutional feature extractors for malaria parasite classification in erythrocyte microscopy," *IEEE Access*, vol. 13, pp. 14520–14532, 2025.
- [18] S. Rajaraman *et al.*, "MalariAI: A label-resilient decoupled framework for universal cell segmentation and explainable stage classification in dense malaria blood smears," arXiv preprint arXiv:2502.01234, 2025.