

An Explainable Deep Learning for Malaria Blood Cell Classification Using DenseNet121 and Grad-CAM

Octavian^{1*}, Imelda Widjaja², Supri Amir³

^{1,3} Department of Information Systems, Hasanuddin University, Makassar, Indonesia ² Faculty of Medicine, Hasanuddin University, Makassar, Indonesia

ARTICLE INFO

Article history:

Initial submission

16-01-2026

Received in revised form

20-01-2026

Accepted 27-01-2026

Available online 27-02-2026

Keywords:

Malaria Detection, Deep Learning, Explainable Artificial Intelligence, Medical Image Analysis

DOI:

<https://doi.org/10.59356/smart-techno.v8i1.199>

ABSTRACT

Malaria diagnosis based on microscopic examination of blood smears is time-consuming and highly dependent on skilled laboratory personnel, which limits its scalability in resource-constrained environments. This study investigated whether an explainable deep learning approach could provide reliable and interpretable malaria blood cell classification using a convolutional neural network based on the DenseNet121 architecture combined with Gradient-weighted Class Activation Mapping to visualize image regions influencing model predictions. Five-fold cross-validation was applied to ensure stable and unbiased performance evaluation. The proposed model achieved a mean classification accuracy of 0.8285 with a standard deviation of 0.0052, demonstrating consistent generalization across folds, while precision, recall, and F1-score values remained balanced between parasitized and uninfected classes. Grad-CAM visualizations applied to representative samples consistently highlighted intracellular regions associated with parasite presence in infected cells and more uniform cytoplasmic regions in uninfected samples, providing qualitative evidence that the model learned biologically meaningful

features. These results indicate that DenseNet121 offers a stable and interpretable solution for malaria blood cell classification when supported by visual explanation, enabling transparent automated screening suitable for smart healthcare and medical informatics applications.

1. INTRODUCTION

Malaria remains one of the most persistent and damaging infectious diseases worldwide. In many regions, particularly in low and middle-income countries, it continues to place a heavy burden on public health systems, local economies and vulnerable populations (Anikeeva et al., 2024). Despite decades of scientific progress, the disease is still responsible for hundreds of millions of infections each year, and early diagnosis remains one of the most effective strategies for reducing its transmission and mortality. However, diagnostic practices have not evolved at the same pace as treatment (Yalley et al., 2024).

Microscopic examination of stained blood smears remains the clinical gold standard for malaria detection. However, this approach is slow, labor-intensive, and highly dependent on the expertise of trained laboratory personnel (Davidson et al., 2021). Even minor variations in staining quality, illumination, or sample preparation can significantly degrade diagnostic accuracy, a problem that is especially severe in rural and resource-limited settings where laboratory conditions are rarely optimal. These constraints have motivated the increasing use of artificial intelligence in automated blood smear analyses. Over the last decade, deep learning, particularly convolutional neural networks (CNNs), has transformed image analysis by learning complex visual representations directly from raw pixel data and achieving state-of-the-art performance across a wide range of medical imaging tasks, from tumor detection to retinal disease screening (Ramos-Briceño et al., 2025). Consequently, CNN-based malaria blood smear classification has emerged as a widely studied and highly effective approach, with numerous reports demonstrating strong performance in distinguishing parasitized from uninfected blood cells (Quan et al., 2020).

Despite these advances, high classification accuracy alone is not sufficient for reliable clinical use, particularly when using convolutional neural networks, such as DenseNet121, which operate as complex black-box models (Laschowski et al., 2022). These networks can produce highly confident predictions without revealing how or why those decisions were made, which represents a critical limitation in medical contexts, where clinicians must understand the basis of a diagnosis before trusting or acting on it. This lack of transparency has driven the rapid development of explainable artificial intelligence (XAI), which aims to make the behavior of deep learning models interpretable by human users. Among XAI techniques, Gradient-weighted Class Activation Mapping (Grad-CAM) has become one of the most widely adopted methods for convolutional neural networks, including DenseNet architectures, by generating heatmaps that indicate which regions of an image contribute most strongly to a model's prediction (Ennab & Mcheick, 2025). In malaria microscopy, these visualizations can reveal whether a DenseNet-based model focuses on parasite bodies within red blood cells or responds to irrelevant background features.

This study addresses this need by systematically evaluating a DenseNet121-based malaria classification model with Grad-CAM-based visual explanations. Rather than proposing a new architecture, the focus is on rigorously measuring the classification performance and analyzing what the model actually learns to see in blood smear images. By integrating cross-validated DenseNet121 classification with Grad-CAM visualization, this study aims to provide a transparent and reproducible assessment of how deep learning models interpret malaria microscopy, contributing to the development of more trustworthy and clinically meaningful AI-based diagnostic systems.

2. LITERATURE REVIEW

Deep learning has become the dominant approach for automated malaria diagnosis from blood smear images, largely because of the success of convolutional neural networks (CNNs) in medical image analysis (Sriporn et al., 2020). A wide range of architectures, including VGG, ResNet, Inception, and DenseNet, have been applied for malaria classification (Alraba'nah & Toghuji, 2024; Loddo et al., 2022; Nakasi et al., 2020).

Although the strong performance of modern convolutional neural networks demonstrates their effectiveness for malaria image classification, recent research has increasingly emphasized that high accuracy alone is insufficient for reliable medical deployment because deep learning models are typically opaque, making it difficult to determine whether predictions are driven by clinically meaningful structures or spurious visual correlations (Ibrahim & Shafiq, 2023). This has motivated the adoption of explainable artificial intelligence (XAI) techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM), which provides visual explanations by highlighting the image regions that most strongly influence a model's output (M et al., 2024). In malaria microscopy, these visualizations can indicate whether the CNNs attend to the parasite regions within red blood cells, thereby offering an intuitive insight into the model behavior. In most existing studies (Rahman et al., 2024; Xiao et al., 2021; Zhang & Ogasawara, 2023), Grad-CAM has primarily been used as a qualitative tool to support model interpretation rather than as a fully quantitative evaluation of attention consistency, and this practice remains common in applied medical imaging research.

Motivated by this perspective, the present study employed DenseNet121 as a balanced and computationally efficient convolutional neural network architecture that offers an effective trade-off between model depth, learning stability, and interpretability, rather than maximizing classification accuracy through deeper variants such as DenseNet169 or DenseNet201 (Lakshmi & Sargunam, 2024). DenseNet121 was selected as a representative baseline model that is widely used in medical image analysis and suitable for visual explanation, as its moderate depth facilitates clearer and more interpretable activation patterns when applying Grad-CAM. Accordingly, this study did not aim to establish new state-of-the-art performance in terms of accuracy; instead, Grad-CAM was applied to representative examples to illustrate how the model formed its predictions and to verify that decision-making was based on

biologically plausible image regions. This approach supports the objective of developing a transparent and practically useful malaria classification system for smart healthcare and medical informatics applications.

3. METHOD

3.1 Research Workflow

To provide a clear overview of the proposed approach, this subsection describes the end-to-end workflow of the explainable malaria classification system. The workflow was designed to integrate data processing, model training, performance evaluation, and visual explanation into a single coherent pipeline, ensuring both predictive reliability and interpretability.

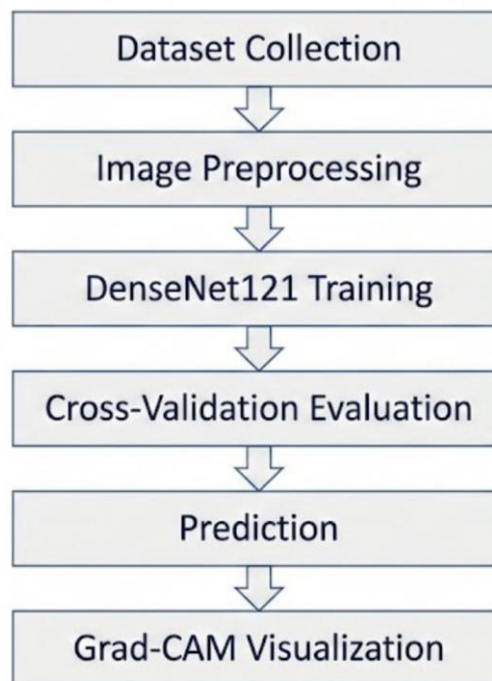


Figure 1. Overall research workflow of the proposed system

Figure 1 illustrates the overall workflow of the proposed explainable malaria classification system, beginning with data acquisition and ending with visual interpretation of model predictions. The process starts with the collection of microscopic blood smear images, which are subsequently preprocessed through resizing and normalization to ensure compatibility with the DenseNet121 input requirements. The preprocessed images are then used to train and evaluate the DenseNet121 model using a five-fold cross-validation strategy, allowing stable and unbiased performance assessment across different data partitions. Following classification, Gradient-weighted Class Activation Mapping is applied to representative test images to generate visual heatmaps that highlight the image regions contributing most strongly to each prediction. This workflow integrates quantitative performance evaluation with qualitative visual explanation, providing a transparent and interpretable framework for automated malaria blood cell classification within smart healthcare and medical informatics applications.

3.2 Data Source and Collection

The dataset used in this study was obtained from a publicly available malaria microscopy repository containing segmented images of red blood cells labeled as either parasitized or uninfected (Rajaraman et al., 2019). The dataset consisted of thousands of high-

resolution blood smear images captured under controlled laboratory conditions, with each image representing a single cell extracted from a thin blood smear slide. The use of a standardized and openly accessible dataset enabled reproducibility and facilitated direct comparison with previous studies that evaluated deep learning models for malaria diagnosis.

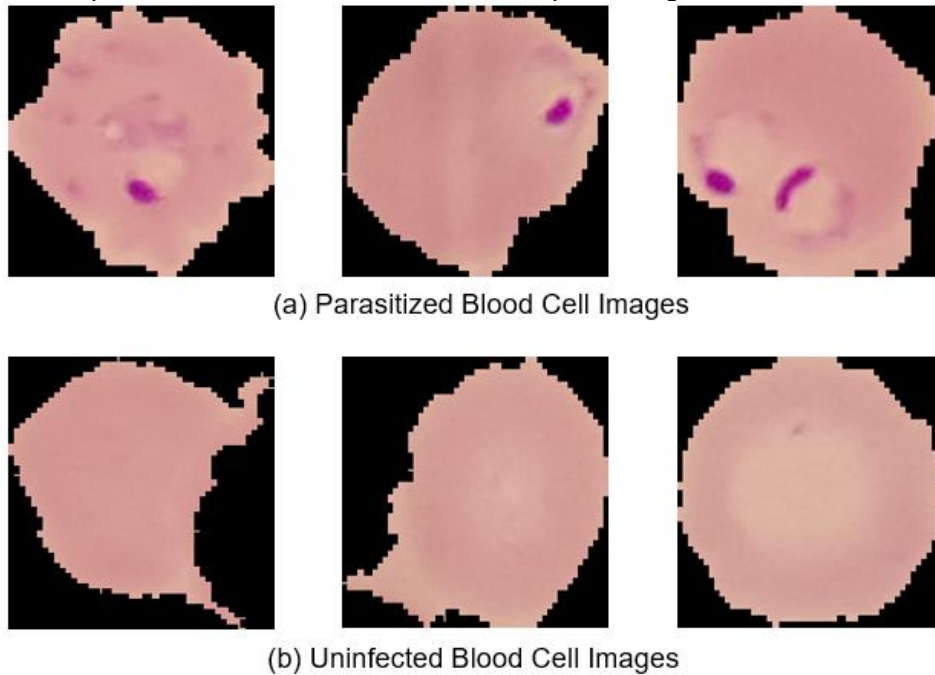


Figure 2. Images of Parasitized and Uninfected Blood Cell

The images were organized into two classes corresponding to infected and healthy cells, and a sample image is shown in **Figure 2**. The dataset consisted of 27,560 segmented blood cell images, with 13,780 images for each class (parasitized and uninfected), ensuring perfect class balance. A five-fold cross-validation strategy was employed to evaluate model performance. In each fold, the dataset was first divided into training and testing subsets using an 80:20 ratio at the class level. The training subset was then further split into training and validation sets using an additional 80:20 ratio. This resulted in 8,818 images used for training, 2,205 images for validation, and 2,756 images for testing per class in each fold. **TABLE 1** provides the detail of dataset distribution in this research

Table 1. Dataset Distribution per Fold

Subset	Parasitized	Uninfected	Total Images
Training	8,818	8,818	17,636
Validation	2,205	2,205	4,410
Testing	2,756	2,756	5,512
Total	13,780	13,780	27,560

To ensure consistency in training and evaluation, all images were sorted and indexed deterministically before model construction. Each image was resized to a fixed resolution of 224×224 pixels and normalized to the range $[0, 1]$ to satisfy the input requirements of the DenseNet121 architecture. No manual feature extraction was applied because the convolutional neural network was designed to learn relevant visual features directly from the pixel data.

3.3 Model Architecture

The proposed model is based on DenseNet121, a deep convolutional neural network characterized by dense connectivity between layers, where each layer receives the feature maps of all preceding layers as input. This design encourages feature reuse and improves

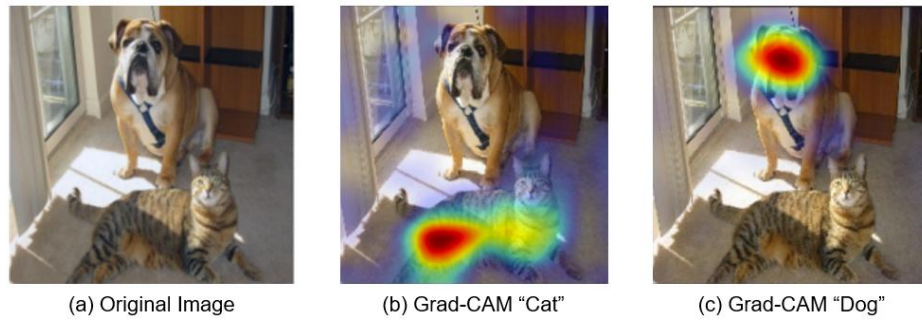


Figure 4. Example of Grad-CAM heatmap (Selvaraju et al., 2017)

Grad-CAM heatmaps were generated for both the parasitized and uninfected classes for each correctly classified test image. For the sigmoid-based binary classifier, the heatmap for the uninfected class was computed by inverting the predicted probabilities. The resulting heatmaps were resized to match the input image resolution and overlaid on the original blood cell images using a color map, allowing for visual inspection of whether the network focused on parasite structures or irrelevant background regions.

4. RESULT AND DISCUSSION

The DenseNet121 model demonstrated a stable learning behavior across all five validation folds. The training and validation curves in **Figure 5** and **Figure 6** showed a smooth increase in accuracy and a consistent decrease in loss, with close alignment between the two, indicating that the network learned generalizable features rather than overfitting to the training data. This stability is particularly important for medical imaging tasks, in which models must perform reliably on unseen data.

Using five-fold cross-validation, the model achieved a mean classification accuracy of 0.8285, with a standard deviation of 0.0052, indicating consistent performance across different data splits. The individual fold accuracies ranged from 0.8224 to 0.8351, demonstrating that the learned feature representations were stable and not overly dependent on specific training subsets. The results in **Table 2** suggest that DenseNet121 offers a dependable and reproducible baseline for malaria classification, balancing effectiveness with computational efficiency.

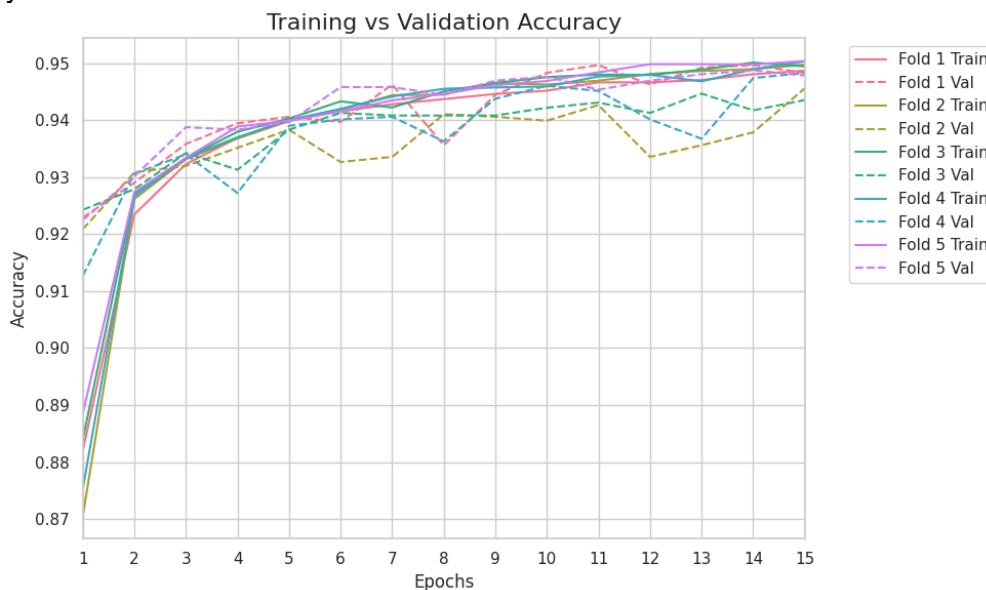


Figure 5. Plot of Training and Validation Accuracy for each Epoch

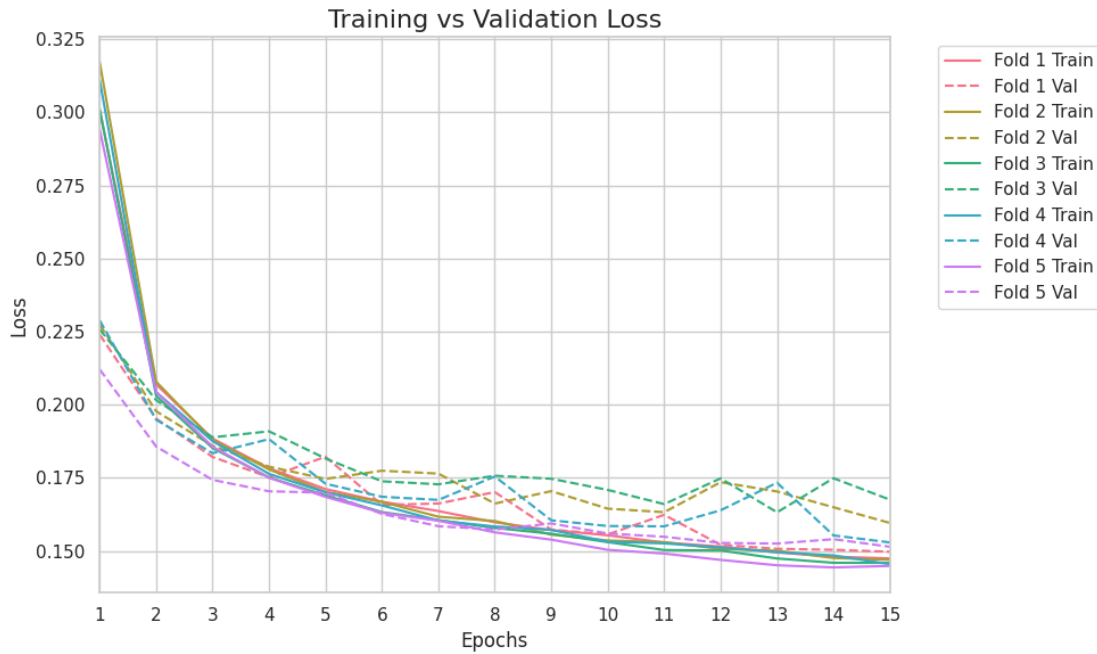


Figure 6. Plot of Training and Validation Loss for each Epoch

Table 2. Accuracy Model for Each Fold

No	Fold	Accuracy
1	Fold 1	0.8351
2	Fold 2	0.8258
3	Fold 3	0.8224
4	Fold 4	0.8247
5	Fold 5	0.8343
6	Mean \pm Std	0.8285 \pm 0.0052

Table 3. Precision, Recall, and F1-Score of Model for each Fold

Fold	Class	Precision	Recall	F1-score
Fold 1	Parasitized	0.85	0.82	0.84
	Uninfected	0.82	0.85	0.84
Fold 2	Parasitized	0.83	0.82	0.82
	Uninfected	0.82	0.83	0.83
Fold 3	Parasitized	0.84	0.80	0.82
	Uninfected	0.81	0.85	0.83
Fold 4	Parasitized	0.84	0.81	0.82
	Uninfected	0.81	0.84	0.83
Fold 5	Parasitized	0.84	0.83	0.83
	Uninfected	0.83	0.84	0.84

The performance demonstrated a well-balanced effectiveness between the parasitized and uninfected classes, which is essential for reliable malaria detection. Throughout all folds in the cross-validation process, the macro-averaged F1-scores consistently ranged from 0.82 to 0.84, indicating stable and robust classification capability. Moreover, the precision and recall values for both classes were closely aligned, reflecting that the model maintained comparable sensitivity and specificity without favoring one class over the other. This equilibrium is particularly important in the clinical context of malaria screening, where minimizing both false negatives, missed infections, and false positives, incorrect diagnoses, is critical to patient outcomes and resource allocation. The absence of systematic bias toward either parasitized

or uninfected samples further reinforces the model's potential for practical deployment in diagnostic-support systems, ensuring equitable performance across diverse cases. These findings underscore the model's suitability as a dependable tool that can assist healthcare professionals by providing consistent and clinically meaningful predictions. The details score can be seen in **Table 3**.

Confusion matrices, **Figure 7**, further illustrate this balanced behavior. In Fold 1, for example, 2298 parasitized and 2297 uninfected cells were correctly classified, with misclassifications distributed relatively evenly. Similar patterns were observed across all folds, reinforcing the conclusion that the classifier maintains consistent decision boundaries between infected and healthy cells.

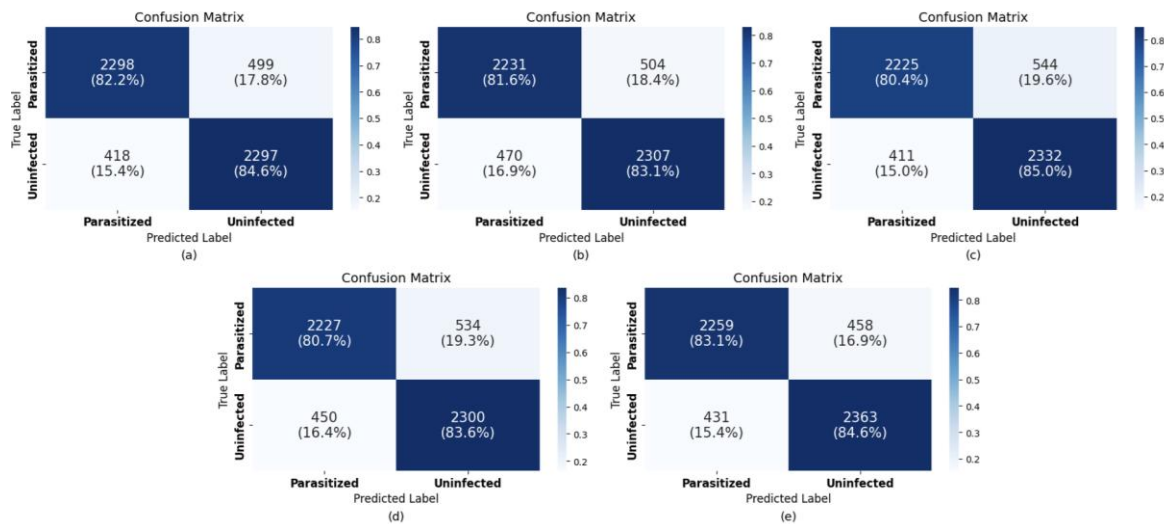


Figure 7. Confusion Matrices Across All Folds

Beyond numerical performance, Grad-CAM visualization provides insight into what the model learns. The heatmaps consistently highlight intracellular regions associated with parasite presence in infected cells, while focusing on more uniform cytoplasmic regions in uninfected samples. These visual patterns align with biological understanding of malaria morphology and indicate that the DenseNet121 model bases its predictions on diagnostically relevant features rather than on background artifacts or staining irregularities.

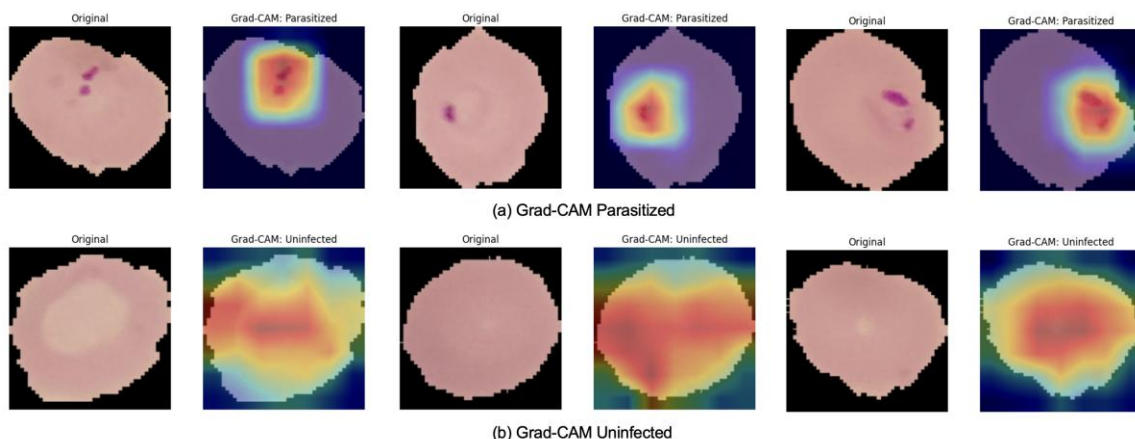


Figure 8. Representative Grad-CAM overlays for parasitized and uninfected cells

Grad-CAM visualizations provide qualitative insights into how the DenseNet121 model predicts parasitized and uninfected blood cell images. For parasitized samples, the Grad-CAM

heatmaps consistently highlighted localized intracellular regions that corresponded to the presence of parasite structures. As shown in **Figure 8(a)**, high-activation areas (indicated by warmer colors) are concentrated around the purple-stained parasite bodies within the red blood cells. This suggests that the model relies on specific morphological cues associated with infection rather than diffuse background information when predicting the parasitized class of an organism.

In contrast, the Grad-CAM results for the uninfected samples exhibited different activation patterns. As illustrated in **Figure 8(b)**, the heatmaps were more broadly distributed across the cytoplasmic region of the cell, with no strong localized focus on specific intracellular structures. This behavior reflects the absence of parasite morphology and indicates that the model captures the overall uniform texture and appearance of healthy blood cells when predicting an uninfected class. The contrast between these two patterns demonstrates that the model differentiates between infected and uninfected cells not only through numerical classification scores but also through distinct visual attention strategies. For parasitized cells, attention is concentrated and localized, whereas for uninfected cells, attention is more diffuse and evenly spread. Although Grad-CAM was applied to representative examples rather than exhaustively across the dataset, the observed visualizations provide intuitive evidence that the model predictions are grounded in biologically plausible image regions.

Overall, these Grad-CAM results support the interpretability of the DenseNet121 model by demonstrating that its decisions are influenced by meaningful cellular features. This qualitative explainability strengthens confidence in the model's behavior and highlights the value of integrating visual interpretation into deep learning-based malaria diagnosis systems, particularly in smart healthcare and medical informatics applications.

5. CONCLUSION

This study examined whether an explainable deep learning approach based on DenseNet121 and Grad-CAM could provide reliable and interpretable malaria blood cell classification. The experimental results demonstrated that the proposed model achieved a mean classification accuracy of 0.8285 with a standard deviation of 0.0052 across five-fold cross-validation, indicating stable and consistent generalization performance. In addition to numerical performance, Grad-CAM visualizations applied to representative samples showed that the model focused on intracellular regions associated with parasite presence in infected cells and more uniform cytoplasmic regions in uninfected cells, providing qualitative evidence that the network learned biologically meaningful features. These findings confirm that a moderately deep convolutional neural network, when combined with visual explanation, can deliver transparent and clinically meaningful predictions without relying on the most complex architectures. Based on these results, developers of smart healthcare and medical informatics systems are encouraged to emphasize interpretability and stability alongside accuracy, while healthcare practitioners and system designers may use explainable models such as DenseNet121 with Grad-CAM to support trustworthy and practical malaria screening solutions, particularly in resource-limited and telemedicine-based settings.

6. ACKNOWLEDGEMENT

The author declares no conflict of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

7. REFERENCES

- Alraba'nah, Y., & Toghuji, W. (2024). A deep learning based architecture for malaria parasite detection. *Bulletin of Electrical Engineering and Informatics*, 13(1), 292–299. <https://doi.org/10.11591/eei.v13i1.5485>
- Anikeeva, O., Hansen, A., Varghese, B., Borg, M., Zhang, Y., Xiang, J., & Bi, P. (2024). The impact of increasing temperatures due to climate change on infectious diseases. *BMJ*, 387, e079343. <https://doi.org/10.1136/bmj-2024-079343>

- Davidson, M. S., Andradi-Brown, C., Yahiya, S., Chmielewski, J., O'Donnell, A. J., Gurung, P., Jeninga, M. D., Prommana, P., Andrew, D. W., Petter, M., Uthaipibull, C., Boyle, M. J., Ashdown, G. W., Dvorin, J. D., Reece, S. E., Wilson, D. W., Cunningham, K. A., Ando, D. Michael., Dimon, M., & Baum, J. (2021). Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks. *Biological Imaging*, 1, e2. <https://doi.org/10.1017/S2633903X21000015>
- Ennab, M., & McHeick, H. (2025). Advancing AI Interpretability in Medical Imaging: A Comparative Analysis of Pixel-Level Interpretability and Grad-CAM Models. *Machine Learning and Knowledge Extraction*, 7(1), 12. <https://doi.org/10.3390/make7010012>
- Ibrahim, R., & Shafiq, M. O. (2023). Explainable Convolutional Neural Networks: A Taxonomy, Review, and Future Directions. *ACM Computing Surveys*, 55(10), 1–37. <https://doi.org/10.1145/3563691>
- Lakshmi, K. S., & Sargunam, B. (2024). Exploration of AI-powered DenseNet121 for effective diabetic retinopathy detection. *International Ophthalmology*, 44(1), 90. <https://doi.org/10.1007/s10792-024-03027-7>
- Laschowski, B., McNally, W., Wong, A., & McPhee, J. (2022). Environment Classification for Robotic Leg Prostheses and Exoskeletons Using Deep Convolutional Neural Networks. *Frontiers in Neurobotics*, 15. <https://doi.org/10.3389/fnbot.2021.730965>
- Loddo, A., Fadda, C., & Di Ruberto, C. (2022). An Empirical Evaluation of Convolutional Networks for Malaria Diagnosis. *Journal of Imaging*, 8(3), 66. <https://doi.org/10.3390/jimaging8030066>
- M, M. M., T. R, M., V, V. K., & Guluwadi, S. (2024). Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50. *BMC Medical Imaging*, 24(1), 107. <https://doi.org/10.1186/s12880-024-01292-7>
- Ma, K., Liu, W., Zhang, K., Duanmu, Z., Wang, Z., & Zuo, W. (2018). End-to-End Blind Image Quality Assessment Using Deep Neural Networks. *IEEE Transactions on Image Processing*, 27(3), 1202–1213. <https://doi.org/10.1109/TIP.2017.2774045>
- Nakasi, R., Mwebaze, E., Zawedde, A., Tusubira, J., Akeru, B., & Maiga, G. (2020). A new approach for microscopic diagnosis of malaria parasites in thick blood smears using pre-trained deep learning models. *SN Applied Sciences*, 2(7), 1255. <https://doi.org/10.1007/s42452-020-3000-0>
- Quan, Q., Wang, J., & Liu, L. (2020). An Effective Convolutional Neural Network for Classifying Red Blood Cells in Malaria Diseases. *Interdisciplinary Sciences: Computational Life Sciences*, 12(2), 217–225. <https://doi.org/10.1007/s12539-020-00367-7>
- Rahman, M. F., Tseng, T.-L. (Bill), Pokojovy, M., McCaffrey, P., Walser, E., Moen, S., Vo, A., & Ho, J. C. (2024). Machine-Learning-Enabled Diagnostics with Improved Visualization of Disease Lesions in Chest X-ray Images. *Diagnostics*, 14(16), 1699. <https://doi.org/10.3390/diagnostics14161699>
- Rajaraman, S., Jaeger, S., & Antani, S. K. (2019). Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *PeerJ*, 7, e6977. <https://doi.org/10.7717/peerj.6977>
- Ramos-Briceño, D. A., Flammia-D'Aleo, A., Fernández-López, G., Carrión-Nessi, F. S., & Forero-Peña, D. A. (2025). Deep learning-based malaria parasite detection: convolutional neural networks model for accurate species identification of *Plasmodium falciparum* and *Plasmodium vivax*. *Scientific Reports*, 15(1), 3746. <https://doi.org/10.1038/s41598-025-87979-5>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV), 618–626. <https://doi.org/10.1109/ICCV.2017.74>
- Sriporn, K., Tsai, C.-F., Tsai, C.-E., & Wang, P. (2020). Analyzing Malaria Disease Using Effective Deep Learning Approach. *Diagnostics*, 10(10), 744. <https://doi.org/10.3390/diagnostics10100744>
- Xiao, M., Zhang, L., Shi, W., Liu, J., He, W., & Jiang, Z. (2021). A visualization method based on the Grad-CAM for medical image segmentation model. 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), 242–247. <https://doi.org/10.1109/EIECS53707.2021.9587953>
- Yalley, A. K., Ocran, J., Cobbinah, J. E., Obodai, E., Yankson, I. K., Kafintu-Kwashie, A. A., Amegatcher, G., Anim-Baidoo, I., Nii-Trebi, N. I., & Prah, D. A. (2024). Advances in Malaria Diagnostic Methods in Resource-Limited Settings: A Systematic Review. *Tropical Medicine and Infectious Disease*, 9(9), 190. <https://doi.org/10.3390/tropicalmed9090190>
- Zhang, H., & Ogasawara, K. (2023). Grad-CAM-Based Explainable Artificial Intelligence Related to Medical Text Processing. *Bioengineering*, 10(9), 1070. <https://doi.org/10.3390/bioengineering10091070>