International Journal of Robotics and Control Systems
Vol. 6, No. 1, 2026, pp. 383-400
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs

A Vision Transformer Architecture for Automated Recognition of Parasitic Types in Microscopic Images

Yuri Pamungkas [a,1,*], Evi Triandini [b,2], Abdul Karim [c,3], Thosporn Sangsawang [d,4]

[a] Department of Medical Technology, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
[b] Department of Information System, Institut Teknologi dan Bisnis STIKOM Bali, Denpasar, 80234, Indonesia
[c] Department of Artificial Intelligence Convergence, Hallym University, Chuncheon, 24252, Republic of Korea
[d] Division of Educational Technology and Communications, Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110, Thailand
[1] yuri@its.ac.id; [2] evi@stikom-bali.ac.id; [3] abdullkarim@korea.ac.kr; [4] sthosporn@rmutt.ac.th
* Corresponding Author

ARTICLE INFO

Article history
Received September 29, 2025
Revised November 01, 2025
Accepted February 01, 2026

Keywords
Vision Transformer;
Parasite Recognition;
Microscopic Images;
Explainable AI;
Medical Image Classification

ABSTRACT

Parasitic infections continue to pose a major global health challenge, with diagnosis still largely dependent on manual microscopic examination. Although CNNs have been applied to automate parasite detection, they are limited in capturing global context, which is crucial for distinguishing subtle morphological differences. To overcome these limitations, this study introduces a Vision Transformer (ViT) architecture for automated recognition of multiple parasite species and host cells in microscopic images. The proposed approach was evaluated on a dataset of 34,298 images across eight classes, including Babesia, Leishmania, Leukocyte, Plasmodium, Red Blood Cells (RBCs), Toxoplasma, Trichomonad, and Trypanosome. Images were preprocessed and augmented before being transformed into patch embeddings and passed through a series of transformer encoding modules employing multi-headed self-attention mechanisms to capture contextual dependencies across the image patches. A classification head produced predictions, while interpretability was examined using Grad-CAM and Score-CAM. Results show that the ViT model achieved excellent performance, with an accuracy of 99.70%, precision of 99.46%, recall of 99.40%, specificity of 99.60%, and F1-score of 99.45%. Confusion matrix analysis confirmed reliable predictions across all classes, and ROC curves yielded AUC values close to 1.0. Visualization demonstrated that the model consistently focused on biologically meaningful features, with Score-CAM offering sharper localization compared to Grad-CAM. In conclusion, the proposed ViT architecture provides a highly accurate and interpretable framework for parasite recognition, demonstrating strong potential to improve diagnostic workflows and support reliable clinical decision-making.

1. Introduction

Parasitic infections remain a significant global concern in the field of public health, particularly in tropical and subtropical regions where numerous protozoan species are persistently prevalent [1]. Accurate and timely recognition of protozoan organisms such as Plasmodium, Toxoplasma, Babesia, Leishmania, Trypanosoma, and Trichomonas species is critical for effective treatment and disease

control [2]. However, conventional microscopic examination depends substantially on the specialized knowledge of trained professionals, is labor-intensive and susceptible to inconsistencies among different observers, leading to potential misdiagnosis [3]. Moreover, in resource-limited settings, the scarcity of skilled personnel further exacerbates diagnostic delays and compromises patient outcomes [4]. A related challenge lies in the differentiation of these parasites from host cells such as RBCs and leukocytes, which often exhibit overlapping morphological characteristics in blood smears [5].

In order to overcome these obstacles, researchers have progressively explored automated diagnostic approaches supported by computer assistance. Deep learning models, particularly CNNs, have been the dominant architecture for microscopic image classification, demonstrating strong performance in malaria and toxoplasmosis detection [6]. Nonetheless, CNNs face constraints in recognizing extended dependencies and broader contextual interactions, which are crucial for distinguishing morphologically similar parasites [7]. Recent progress in deep neural network methodologies, with a notable emphasis on the Vision Transformer (ViT) model, have demonstrated outstanding performance across multiple computer vision applications by leveraging self-attention modules that facilitate superior modeling of global relationships compared to CNNs [8].

State-of-the-art studies have applied CNNs and hybrid models to specific parasitic infections, often focusing on binary or limited multi-class classification problems. For instance, CNN-based systems have achieved notable accuracy in differentiating Plasmodium species, while other works have employed transfer learning for Leishmania or Toxoplasma recognition [9]. However, only a few approaches address the simultaneous classification of multiple parasites alongside host cells, a scenario that better reflects real clinical conditions where co-existence and morphological overlap are common [10]. Moreover, the application of transformer-based architectures for this multi-class parasitological task remains underexplored. The novelty of this research lies in employing a Vision Transformer framework to perform automated recognition across a broad spectrum of parasites (Plasmodium, Toxoplasma, Babesia, Leishmania, Trypanosome, Trichomonad) and host cells (RBCs and leukocytes) using microscopic images. By exploiting the global self-attention mechanism of ViT, the model is designed to better capture subtle morphological features and contextual relationships that traditional CNNs may overlook.

The contribution of this research is twofold. First, we propose a comprehensive multi-class classification framework using Vision Transformer for simultaneous recognition of parasites and host cells, thereby advancing automated diagnostic systems toward real-world applicability. Second, the study provides a comparative evaluation against state-of-the-art CNN models, highlighting the strengths of ViT in handling complex parasitological image data. Ultimately, this work aims to support more accurate, scalable, and reliable parasite diagnosis, especially in settings where expert microscopy is limited.

2. Related Works

Automated recognition of parasites in microscopic images has been extensively studied using deep learning methods, particularly CNNs. Early research primarily focused on lightweight detection models. For example, Xu et al. proposed YAC-Net, a lightweight CNN-based model for parasite egg detection using the ICIP 2022 dataset, which achieved high precision (97.8%) and recall (97.7%) through architectural modifications of YOLOv5n [11]. Other studies emphasized malaria detection from thick blood smear images. Oliveira et al. employed pixel classifiers based on HSV components combined with a CNN to classify small and large parasites, reporting an accuracy of 90% and specificity of 96% [12]. Similarly, Boit et al. designed a custom CNN architecture using the NIH malaria dataset, which reached an accuracy of 97.68% and precision of 98.88% [15].

Beyond traditional CNNs, ensemble and transfer learning strategies have been explored. Bhuiyan et al. developed an ensemble model combining VGG16, VGG19, and DenseNet201, which outperformed single networks with an accuracy of 97.92% on NIH cell images [13]. Alassaf et al. applied deep transfer learning with Res2Net and Differential Evolution optimization, achieving

accuracy above 95% [17]. Kundu et al. further enhanced performance using hyper-parameter tuned deep models combining VGG19 feature extraction with CNN-LSTM classifiers, yielding 91% accuracy [18]. To address variations in data heterogeneity, Chaharou et al. introduced an image cropping preprocessing technique to improve classification across diverse patient samples, where DenseNet achieved 97.5% accuracy [16]. For red blood cell (RBC) morphology classification, Khan et al. applied RCNNs with noise filtering, achieving 91% accuracy on the Kaggle dataset [14], while Muhammad et al. classified rouleaux formations with multiple CNN architectures, where DenseNet121 outperformed others with 99% accuracy [19].

Recently, semi-supervised learning approaches have gained attention to reduce the heavy reliance on labeled data. Ha et al. proposed a SSGL framework that integrates CNN feature extraction with graph convolutional networks. Their model achieved 91.75% accuracy and 97.25% specificity while using only 20% labeled data, demonstrating the effectiveness of graph-based representation learning in parasite recognition [20]. In summary, previous works demonstrate significant progress in parasite classification using CNN-based, ensemble, transfer learning, and semi-supervised frameworks. However, most approaches rely heavily on convolutional architectures, which may be limited in capturing long-range dependencies and global contextual information in parasite morphology. This gap provides motivation for exploring transformer-based models, particularly Vision Transformers, which have shown strong potential in modeling global relationships in complex image classification tasks.

3. Methodology

3.1. Parasite Dataset

In this study, we employed a carefully compiled dataset comprising 34,298 microscopic images of parasites and host cells, captured at magnifications of 400× and 1000× [21]. The dataset encompasses multiple classes of clinically relevant parasites, as well as red and white blood cells to provide a broad and inclusive depiction of biological diversity. The distribution of images across the different categories is as follows: 843 images of Plasmodium, 6,691 images of Toxoplasma, 1,173 images of Babesia, 2,701 samples depicting Leishmania, 2,385 samples of Trypanosoma, and 10,134 samples of Trichomonas. Additionally, the dataset includes 8,995 instances of red blood cells (RBCs) and 1,376 instances of leukocytes. The inclusion of both parasite types and host cells allows the dataset to simulate realistic diagnostic conditions where morphological similarities can pose a challenge for automated recognition. With its diverse and balanced composition, this dataset is particularly well-suited for computer vision applications including image classification, feature extraction, and the creation of automated diagnostic models. Parasite dataset shown in Fig. 1.

3.2. Preprocessing

The preprocessing pipeline for the parasite dataset is essential in improving model effectiveness by ensuring that the image data are properly conditioned for training purposes [22]. The preprocessing steps follow a systematic approach designed to optimize the dataset for computer vision tasks like image classification [23]. The sequence of preprocessing includes resizing, cropping, augmentation, and normalization to improve the model's capacity to adapt and manage variability within the input samples [24]. The initial step involves resizing or cropping the images to a standard size, which ensures that all input images are consistent and can be introduced into the neural architecture without issue [25]. This is crucial, as different parasite species may vary in scale and resolution, and resizing the images helps in normalizing these discrepancies [26]. After resizing, random cropping (RCrop) is applied to further enhance the randomness in the dataset, creating different views of the images and preventing overfitting [27].

Next, the images undergo random flipping as part of the augmentation process. Flipping horizontally simulates different orientations of the parasite images [28], making the model robust to varying angles and orientations in real-world microscopic images. This step is particularly important in microscopy, where parasites may appear in various orientations depending on the slide preparation

and camera angle. Color jitter is then applied to introduce variability in the color distribution of the images [29]. This augmentation technique changes the luminosity, contrast levels, color intensity, and tonal variations of the images, making the model less affected by illumination differences or staining inconsistencies in the dataset [30]. These modifications simulate different image capturing conditions and make the model more robust to natural changes in image quality. After augmentation, the images are converted to tensors using the ToTensor operation. This converts the images into a numerical representation that can be interpreted by deep learning algorithms [31]. The concluding stage of the preprocessing workflow consists of normalization, where pixel values are scaled to a standard range. This contributes to maintaining stability during the training phase and ensures that the model captures the essential characteristics from the data rather than being influenced by the varying pixel intensity ranges across images [32]. Denormalization can also be applied when necessary, particularly when visualizing the results or converting outputs back to their original scale [33].

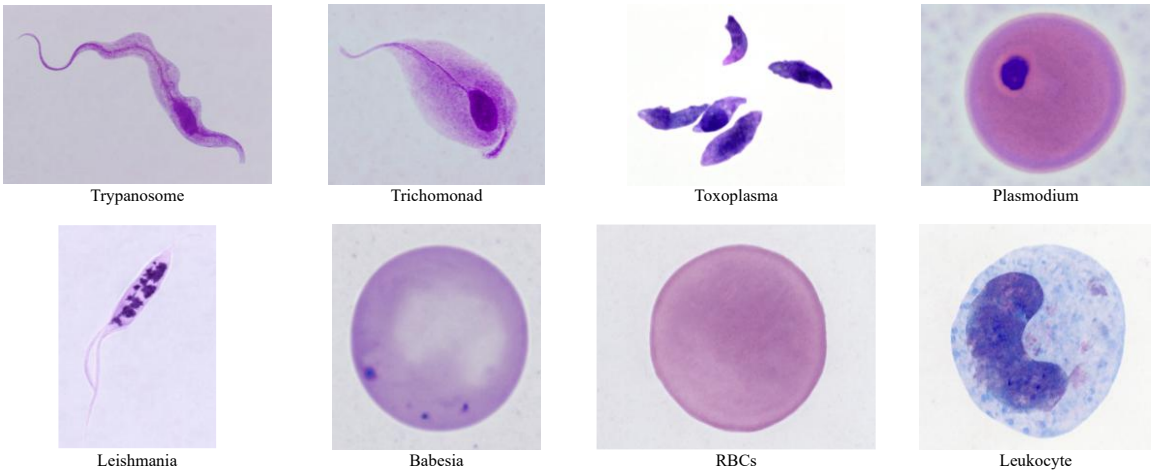


Fig. 1. Parasite dataset

3.3. Vision Transformer (ViT) Architecture

The Vision Transformer (ViT) adapts the transformer framework, originally developed for NLP, to image classification tasks [34]. Unlike CNNs that capture local patterns through convolutional filters, ViT processes an image as a sequence of patches and leverages global self-attention to model long-range dependencies [35]. This capability is particularly important in microscopic parasite recognition, where morphological cues may be subtle and distributed across the cell structure [36]. Proposed vision transformer (ViT) architecture shown in Fig. 2.

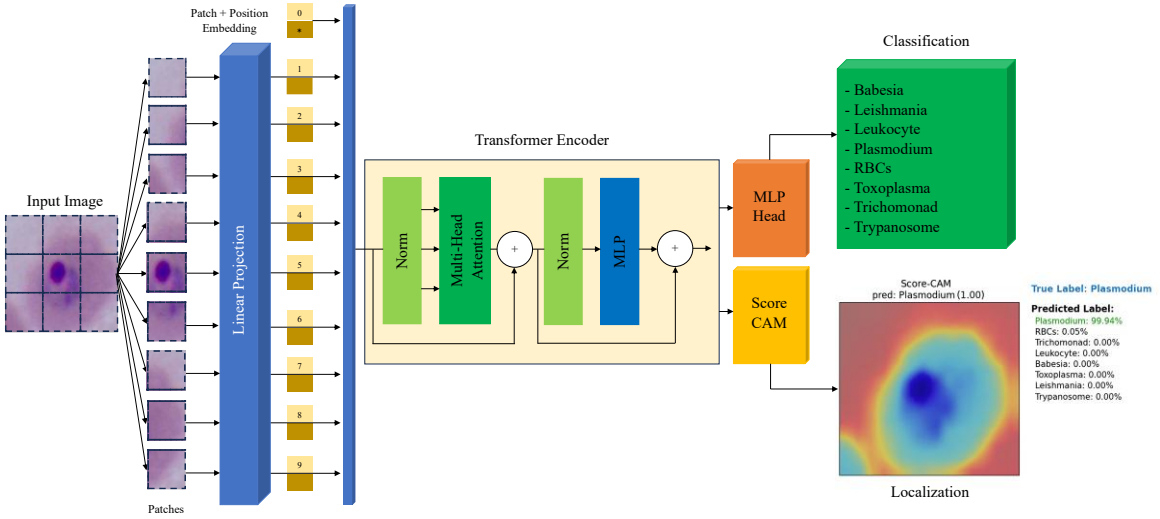


Fig. 2. Proposed vision transformer (ViT) architecture

3.3.1. Patch Embedding

For an input microscopic $x \in R^{H \times W \times C}$, where H denotes the height, W the width, and C the number of channels, the image is segmented into distinct, non-overlapping patches of dimensions $P \times P$. The number of patches is calculated as:

$$N = \frac{H \times W}{P^2} \tag{1}$$

Every patch is reshaped into a one-dimensional vector $x_p \in R^{P^2 \cdot C}$. A linear projection is then applied:

$$z_0 = \left[x_p^1 E; x_p^2 E; \ldots; x_p^N E\right] + E_{pos} \tag{2}$$

where $E \in R^{(P^2 \cdot C) \times D}$ represents the embedding matrix, D corresponds to the dimensionality of the embeddings, and E_{pos} denotes positional embeddings.

3.3.2. Transformer Encoder

Each encoder block consists of a module based on multi-head self-attention (MSA) and a feed-forward neural network composed of multiple layers, also known as a multi-layer perceptron (MLP) [37], with residual connections and a normalization mechanism.

$$z_l' = \text{MSA}\left(\text{Norm}(z_{l-1}')\right) + z_{l-1} \tag{3}$$

$$z_l = \text{MLP}\left(\text{Norm}(z_l')\right) + z_l' \tag{4}$$

where l is the encoder block index.

3.3.3. Multi-Head Self-Attention (MSA)

The self-attention mechanism computes pairwise relationships between image patches [38]. For an input X:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V \tag{5}$$

The scaled dot-product attention is:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

For multiple attention heads:

$$\text{MSA}(X) = [\text{head}_1; \ldots; \text{head}_h]W^O \tag{7}$$

$$\text{head}_i = \text{Attention}\left(XW_i^Q, XW_i^K, XW_i^V\right) \tag{8}$$

where h is the number of heads and $W^O \in R^{hd_v \times D}$.

3.3.4. Multi-Layer Perceptron (MLP)

Each encoder block includes an MLP defined as:

$$\text{MLP}(x) = \text{GELU}(xW_1 + b_1)W_2 + b_2 \tag{9}$$

with learnable weights $W_1 \in R^{D \times d_{ff}}$ and $W_2 \in R^{d_{ff} \times D}$, where d_{ff} is the feed-forward dimension.

3.3.5. Classification Head

After L encoder layers, the output is passed to an MLP classification head:

$$y = \text{Softmax}(z_L W_c + b_c) \quad (10)$$

where $W_c \in R^{D \times K}$, $b_c \in R^K$, and K is the number of classes (Babesia, Leishmania, Leukocyte, Plasmodium, RBCs, Toxoplasma, Trichomonad, Trypanosome).

3.3.6. Explainability with Score-CAM

To improve interpretability, Score-CAM can be applied to emphasize the distinct regions inside the original input image that exerted the most substantial impact on the model's output decision [39]. This becomes especially valuable in medical contexts, as it promotes transparency and builds confidence in computer-assisted diagnostic frameworks [40].

3.4. Metrics Evaluation

To assess the performance capability of the suggested Vision Transformer (ViT) framework in recognizing multiple parasite types from microscopic images, we applied a set of well-established performance metrics. These metrics were chosen because they provide not only an overall picture of how well the model performs, but also more nuanced insights into its strengths and limitations in differentiating infected cells from those that are uninfected. The first metric considered is accuracy, which indicates the ratio of accurately categorized instances within the whole dataset. Accuracy is often reported as the most intuitive measure, since it directly indicates the percentage of correct predictions [41]. However, in medical imaging tasks such as parasite recognition, accuracy alone can be insufficient, especially when there is an imbalance between classes. For example, if one parasite type is underrepresented in the dataset, a model could achieve high accuracy while still performing poorly on that specific class.

To address this limitation, we also examined precision and sensitivity. Precision tells us the ratio of instances identified as positive by the model that genuinely correspond to actual positive cases [42]. In other words, in situations where the model indicates the existence of a parasite like Plasmodium, precision reflects the reliability of that prediction. Conversely, sensitivity quantifies the fraction of authentic positive instances accurately detected by the model [43]. In clinical practice, sensitivity is especially important because failing to detect an infection could result in delayed or missed treatment. Specificity complements sensitivity by measuring the degree to which the model accurately detects negative cases [44]. For instance, high specificity indicates that normal red blood cells or leukocytes are not mistakenly classified as parasites. Together, sensitivity and specificity provide a balanced view of the diagnostic capability of the model, reflecting its competence in identifying actual infections while minimizing erroneous detections.

The F1-score merges precision and recall into one unified measurement, calculated as their harmonic mean [45]. This is particularly useful in multi-class problems like parasite classification, where the model needs to balance the risk of both false-positive and false-negative outcomes at the same time. An elevated F1-score signifies that the network achieves both reliable detection and comprehensive coverage of true positives. Finally, the ROC curve together with the AUC were applied to assess the network's discriminative capability. The ROC curve plots sensitivity against the proportion of incorrect positive classifications across varying decision boundaries, offering an illustrative depiction of the balance between correctly detecting parasites and avoiding false positives [46]. An AUC score approaching 1.0 indicates that the model demonstrates high discriminative strength and is capable of clearly distinguishing between various parasite categories and host cell types.

4. Results and Discussion

In this section, we report the experimental findings derived from utilizing the proposed ViT framework for the automated identification of parasitic types in microscopic images. The evaluation focused on eight distinct classes, namely Babesia, Leishmania, Leukocyte, Plasmodium, RBCs, Toxoplasma, Trichomonad, and Trypanosome. The experiments were structured to evaluate the

effectiveness of the ViT model in distinguishing between parasite species and host cells, while also highlighting its robustness in handling morphological variations across samples. The model underwent training and testing on an extensive dataset with meticulously implemented preprocessing and augmentation techniques to ensure generalization. Performance evaluation was carried out using standard classification metrics such as accuracy, precision, recall (sensitivity), specificity, F1-score, and the ROC curve's AUC. These performance measures deliver a holistic evaluation of the overall predictive performance as well as the diagnostic dependability for each class in the proposed method.

The curves for training and validation shown in Fig. 3 depict the training dynamics exhibited by the proposed ViT architecture over 30 epochs. Both performance accuracy and error rate are reported across both the training and evaluation subsets, providing insight into the network's capability to attain convergence and demonstrate robust generalization performance [47]. At the initial phase of the training process, the model exhibits a swift improvement in the accuracy performance observed on both training and validation phases, rising from approximately 83% at epoch 1 to above 97% by epoch 5. This steep improvement during the initial phase demonstrates that the ViT model rapidly learned to extract distinguishing characteristics of parasites and host cells. After the initial rapid growth, the accuracy continues to improve gradually and stabilizes close to 100% by epoch 20 across the learning and evaluation datasets. The near overlap of the two accuracy curves implies that the model generalizes effectively while avoiding major overfitting.

In terms of loss, the losses from the training and evaluation subsets steadily decline throughout the progression of epochs, confirming that the optimization process is effective. During the initial epochs, the training loss experiences a steep reduction and then settles at minimal values approaching zero. The loss on the validation set, while initially fluctuating slightly in the early epochs, also decreases and stabilizes at similarly low levels as training progresses. This suggests that the network not only adapts effectively to the learning dataset while also preserving strong predictive reliability on unseen validation samples. The combination of steadily increasing accuracy and consistently decreasing loss demonstrates that the ViT architecture is highly effective for the given dataset. Importantly, the lack of separation between the learning and evaluation curves suggests that the network effectively prevents overfitting despite its high capacity [48], which can be ascribed to the implementation of preprocessing techniques, data augmentation methods, and intrinsic regularization mechanisms within the transformer architecture [49]. Overall, the results confirm that the proposed model demonstrates outstanding convergence characteristics and strong generalization capability, making it a reliable tool for automated recognition of parasitic types in microscopic images.

To further examine the classification behavior of the model across different parasite types, a confusion matrix was generated. The confusion matrix offers a comprehensive depiction of the correspondence between actual and predicted labels, allowing for a more profound insight into the model's performance at the class-specific level [50]. This allows researchers to identify not only the ability of the model to accurately categorize different parasite species but also the specific categories where misclassifications may occur [51]. This type of evaluation holds particular importance in multi-class medical imaging tasks [52], where some parasites may share similar morphological features with host cells, leading to potential diagnostic challenges. Beyond serving as a performance summary, the confusion matrix also emphasizes the trade-off between false positives and false negatives, which are both vital for clinical decision-making [53]. False negatives may cause undetected infections, whereas false positives could trigger avoidable treatments or additional invasive interventions [54]. Therefore, minimizing these errors is critical to guarantee that the automated system is reliable in practical diagnostic environments [55]. In this research, the confusion matrix confirms the overall resilience of the ViT-based classifier while also offering practical understanding of its suitability for everyday microscopic diagnostic use. The confusion matrix for the proposed model is presented in Fig. 4.

The evaluation of the proposed ViT model demonstrates consistently high performance across all eight classes of parasites and host cells. As shown in the metrics per class (Fig. 5), the model achieved near-perfect accuracy, precision, sensitivity, specificity, and F1-scores for most categories, underscoring its robustness in distinguishing between morphologically diverse cell types under

microscopic imaging. For Babesia, Leishmania, Leukocyte, RBCs, Toxoplasma, Trichomonad, and Trypanosome, the model obtained accuracy values close to or above 0.997, with precision, sensitivity, specificity, and F1-scores nearly reaching 1.000. This indicates that the model was able to correctly classify almost all true positive cases while avoiding false positives, reflecting its ability to capture discriminative morphological features of these parasites. The particularly high specificity (approaching 1.000 across all classes) further confirms that normal host cells were not misclassified as parasites, which is critical in clinical diagnostic applications.

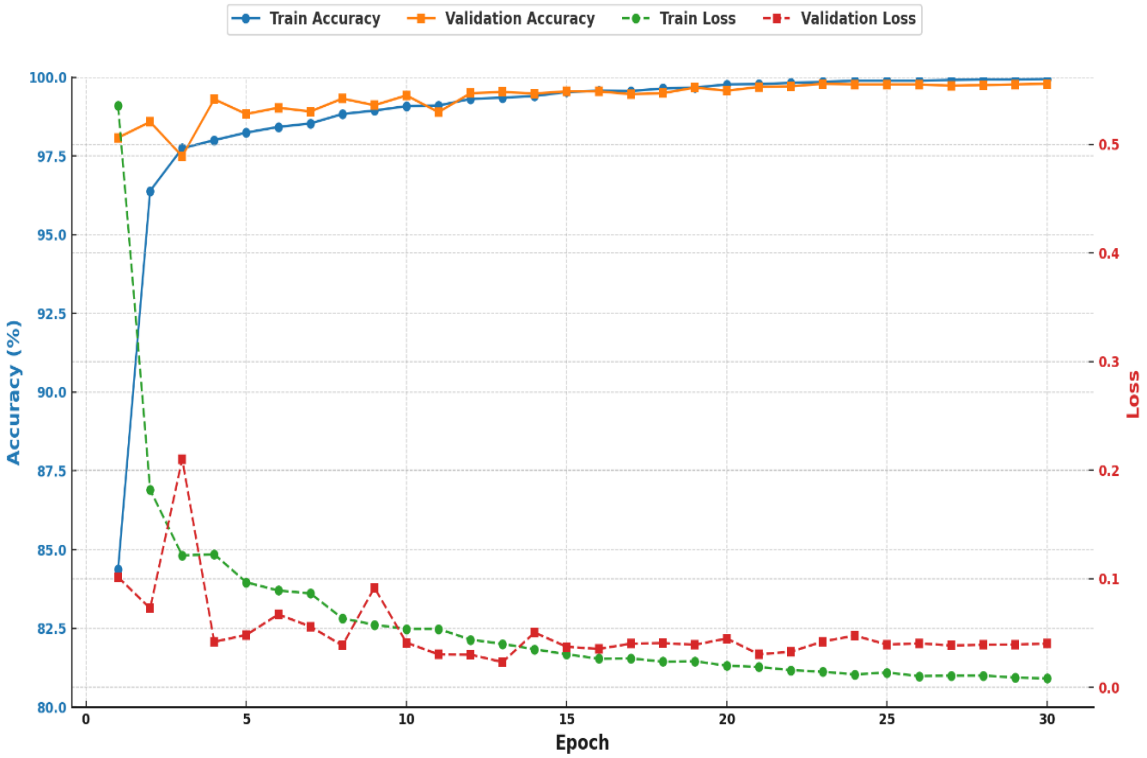


Fig. 3. Training and validation curves

The performance for Plasmodium (although still strong) was relatively lower compared to the other classes. The model achieved an accuracy of 0.997 but showed reduced sensitivity (0.960) and F1-score (0.968). This suggests that a small portion of Plasmodium-infected cells was misclassified, potentially due to their morphological similarity with other cell types, such as red blood cells, or variations in staining and image quality. Nevertheless, the high specificity (0.999) for this class indicates that false positives were rare, ensuring reliable exclusion of non-infected cells. Taken together, these results confirm that the ViT architecture can effectively generalize across multiple parasite classes while maintaining clinical reliability. The near-perfect balance of precision and sensitivity in most classes highlights the model's capability to minimize both false positives and false negatives. Importantly, the slightly lower sensitivity observed for Plasmodium highlights an area for further improvement, potentially through targeted augmentation or inclusion of additional representative samples in the dataset.

The ROC curves shown in Fig. 6 illustrate the discriminative capability of the proposed ViT model across all parasite and host cell classes. The results indicate near-perfect performance, with most classes achieving an AUC of 1.000, including Babesia, Leishmania, Leukocyte, and Toxoplasma. Slightly lower, but still excellent, values were observed for Plasmodium (AUC = 0.980), RBCs (AUC = 0.998), Trichomonad (AUC = 0.999), and Trypanosome (AUC = 0.999). The macro-average AUC of 0.997 further confirms the strong overall generalization ability of the model. These results demonstrate that the ViT-based classifier can effectively distinguish between true positive and false positive cases, reinforcing its reliability and clinical applicability in automated parasite recognition.

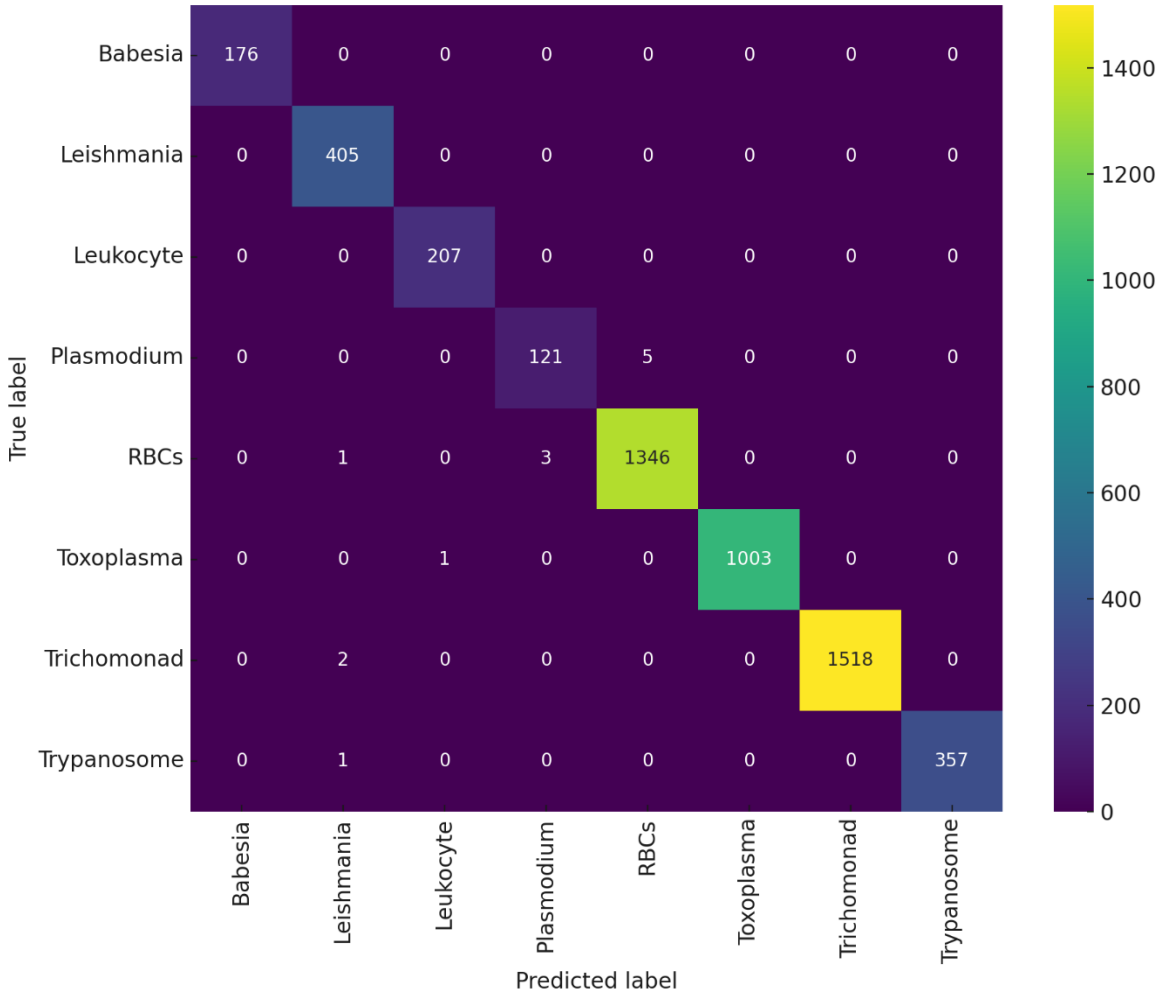


Fig. 4. Confusion matrix

To complement the quantitative evaluation metrics, we further analyzed the interpretability of the proposed ViT model by applying Grad-CAM. Grad-CAM delivers visual interpretations by emphasizing the areas within an input image that most significantly influence the model's output decision [56]. This approach enables validation of whether the model concentrates on biologically meaningful regions [57], such as parasite structures, rather than irrelevant background features. In medical image analysis, such interpretability plays a vital role in fostering trust in computer-assisted diagnostic systems [58]. By visualizing activation maps, clinicians and researchers can better understand the model's decision-making pathway and verify that the classification is grounded in biologically significant morphological attributes [59]. The Grad-CAM results also help to identify potential sources of misclassification, offering insights for subsequent refinement of the model [60]. Overall metrics shown in Fig. 7. Fig. 8 present the Grad-CAM visualizations for representative samples from different parasite classes, illustrating how the ViT-based classifier attends to discriminative regions in microscopic images during the recognition process.

The Grad-CAM and Score-CAM visualizations (Fig. 8) provide valuable insights into how the ViT identifies parasites and host cells. Instead of only reporting accuracy values, these maps show the portions of the image that the model regards as most critical for determining the classification. This holds particular significance in medical imaging, as clinicians require confidence that an automated system emphasizes biologically meaningful features instead of being distracted by non-essential background artifacts [61]. For Babesia, the Score-CAM heatmap showed strong and well-localized attention precisely at the parasite within the red blood cell, a region consistent with what experts would

examine under the microscope. A similar pattern was observed in Leishmania, where the model concentrated on the parasite's elongated body and flagellum. In both cases, the highlighted regions were sharp and biologically convincing, suggesting that the model can robustly capture morphological cues in parasites with distinct shapes.

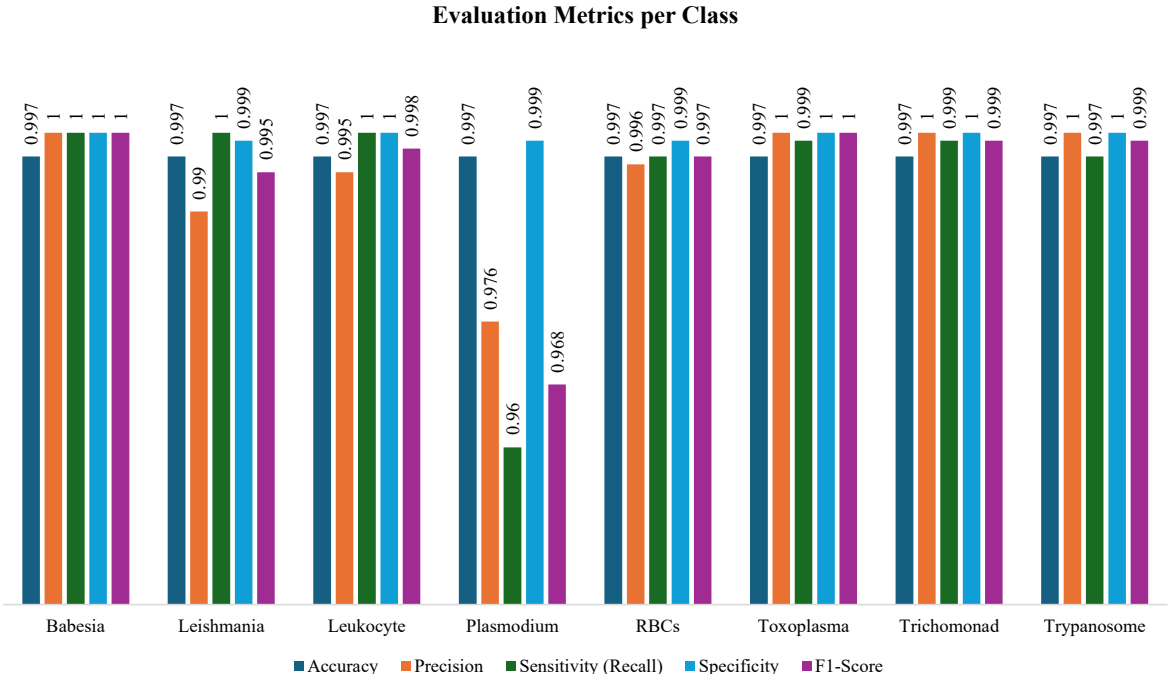


Fig. 5. Evaluation metrics per class of parasitic types classification

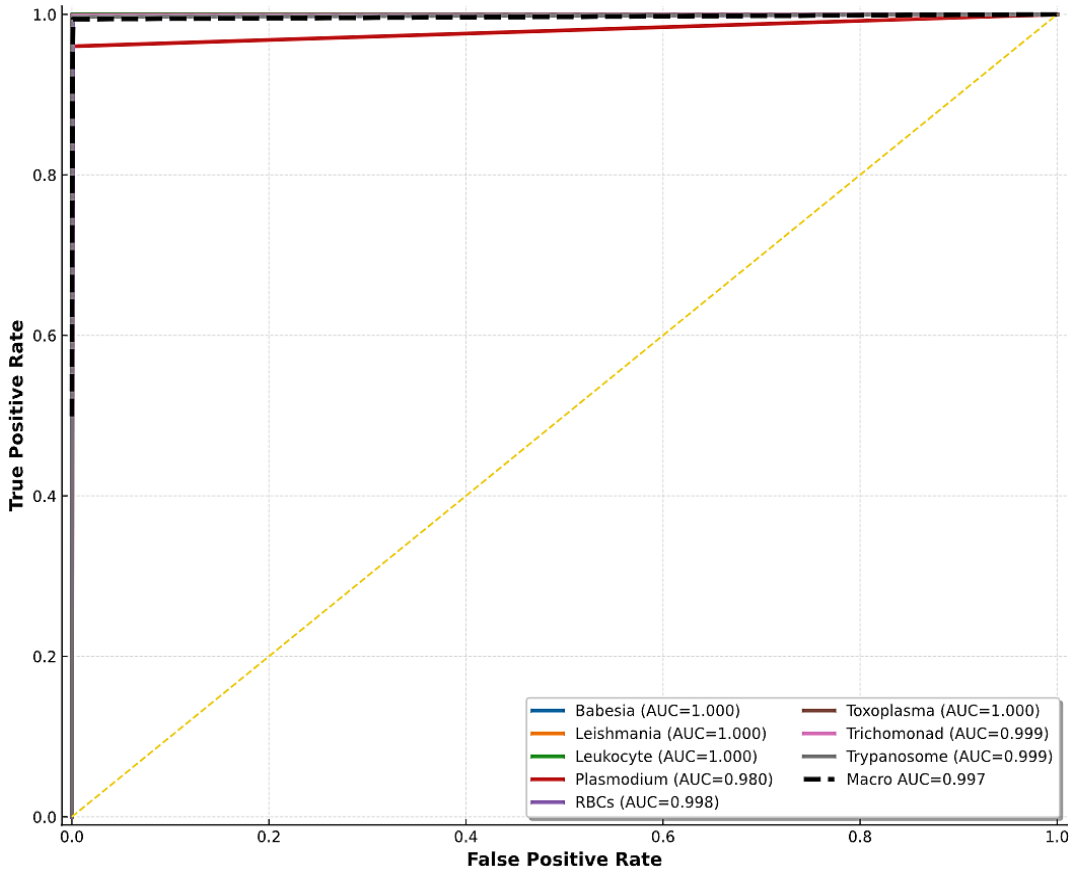


Fig. 6. ROC curves

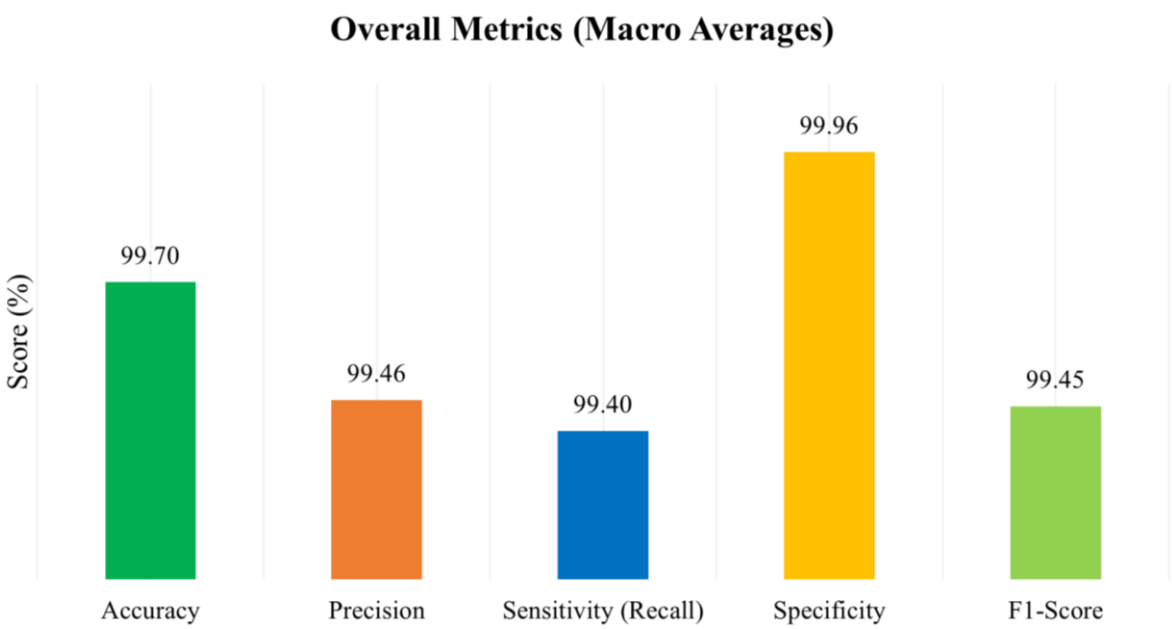


Fig. 7. Overall metrics

In contrast, Plasmodium presented slightly weaker localization. Although the model achieved high classification confidence, the heatmaps were more diffuse compared to Babesia and Leishmania. This could be due to the smaller size and more subtle appearance of Plasmodium within erythrocytes, making it harder for the model to pinpoint the exact parasite boundaries. Nevertheless, the Score-CAM still emphasized the parasite's central body, which confirms that the decision-making process remains biologically relevant. For Trypanosome and Trichomonad, the model produced focused activations along their elongated or oval shapes, successfully capturing their distinctive morphology. Interestingly, these two classes showed clearer and more concentrated attention compared to Plasmodium, again suggesting that larger and more structurally distinct parasites are easier for the ViT to localize. In the case of Toxoplasma, the highlighted region covered most of the parasite body, demonstrating strong agreement between the model's focus and expected biological features.

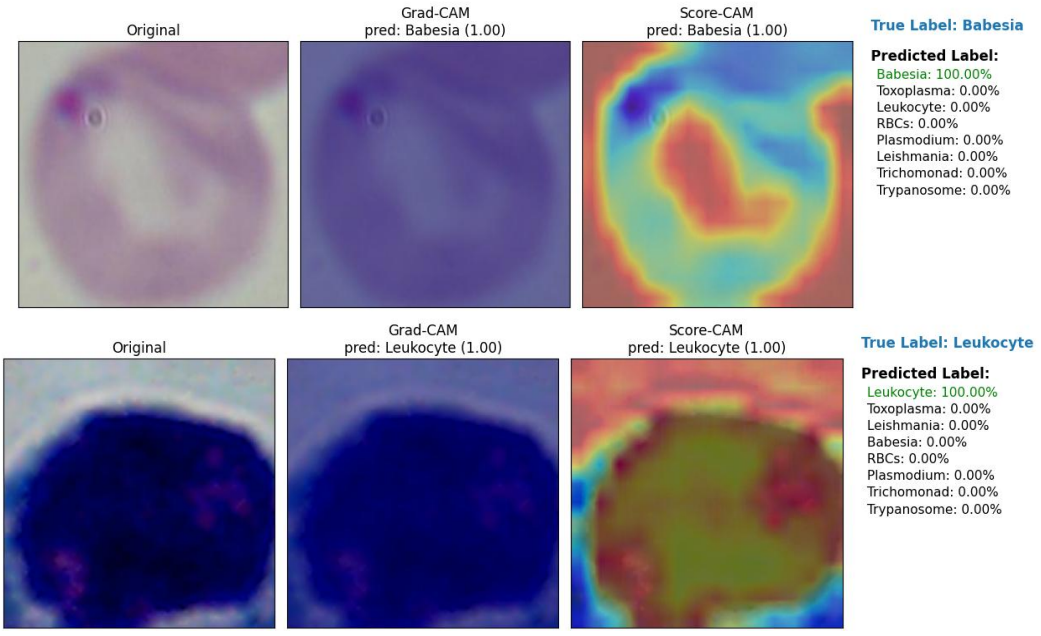
Finally, for Red Blood Cells (RBCs) and Leukocytes, the attention maps revealed that the model correctly focused on the uniform shape of RBCs and the dense nuclear region of leukocytes. These results confirm that the ViT is capable of distinguishing host cells from parasites, minimizing the risk of false positives in practical diagnostic settings. Taken together, these findings indicate that the ViT model tends to produce stronger and sharper localization for parasites with distinctive morphology, such as Babesia, Leishmania, and Trypanosome, while classes like Plasmodium show slightly broader and weaker activations due to their subtle and less differentiated structures. This comparative analysis highlights both the strengths and the remaining challenges for automated parasite recognition and underscores the importance of interpretability methods such as Grad-CAM and Score-CAM in validating deep learning models for medical use.

In addition, Table 1 presents a comparison of our proposed ViT method alongside various cutting-edge approaches for parasite identification. Earlier studies mainly relied on CNN-based architectures, transfer learning, or ensemble models. For instance, Xu et al. [11] developed a lightweight YOLOv5n variant (YAC-Net) with strong performance (F1 = 0.98), while Oliveira et al. [12] applied a CNN to thick smear Plasmodium vivax images but reported a lower accuracy of 90%. Other works, such as Bhuiyan et al. [13] with ensemble learning and Alassaf et al. [17] with transfer learning, achieved accuracies between 95% and 98%. More recent methods have focused on data preprocessing methodologies [16], optimization of hyper-parameters [18], and classification approaches grounded in morphological features [19], with accuracies often above 97%, while semi-supervised approaches like Ha et al. [20] demonstrated strong AUC values but were still limited in handling multiple parasite types simultaneously. In comparison, our ViT model achieved the highest

overall performance across metrics (Acc = 99.70%, Prec = 99.46%, Rec = 99.40%, Spec = 99.60%, F1 = 99.45%), showing that leveraging global attention mechanisms enables more robust classification of diverse parasites and host cells than previous CNN-centric approaches. Importantly, to the best of our awareness, this study is among the earliest to extensively utilize a Vision Transformer framework for identifying multiple parasite species and host cells in microscopic images, offering both superior performance and stronger interpretability compared to prior work.

Table 1. Comparison of our work with some state-of-the-art study techniques

Authors & Year	Objective	Dataset	Method	Results
Xu et al. (2024) [11]	Parasite egg detection (lightweight)	ICIP 2022 (13,200 imgs)	YOLOv5n + AFPN + C2f (YAC-Net)	Prec = 97.8, Rec = 97.7, F1 = 0.98
Oliveira et al. (2022) [12]	Malaria in thick smear	676 P. vivax imgs	CNN (34 layers)	Acc = 90, Sens = 86, Spec = 96
Bhuiyan et al. (2023) [13]	Ensemble malaria detection	NIH (27,558 RBC imgs)	VGG16/19, DenseNet201, Ensemble	Acc/Prec/Rec/F1 = 97.9
Khan et al. (2024) [14]	RBC detection & classification	Kaggle (12,500 imgs)	RCNN	Acc = 91
Boit et al. (2024) [15]	Malaria parasite detection	NIH (27,558 imgs)	Custom CNN	Acc = 97.68, Prec = 98.88
Chaharou et al. (2024) [16]	Malaria, cropping preprocessing	33,007 imgs	CNN, DenseNet, LeNet-5	Acc = 97.5
Alassaf et al. (2022) [17]	Transfer learning malaria	NIH (27,558 imgs)	Res2Net + DE, KNN	Acc = 95.9, Sens = 95.8, Spec = 96.0
Kundu et al. (2023) [18]	Hyper-tuned DL malaria	NIH (27,558 imgs)	VGG19 + CNN-LSTM	Acc = 91, Prec = 89, Rec = 93
Muhammad et al. (2025) [19]	RBC morphology incl. rouleaux	24,712 smear imgs	CNN, Xception, ResNet, DenseNet, EfficientNet	Acc = 99
Ha et al. (2023) [20]	Semi-supervised graph learning parasites	5,758 Plasmodium, 5,878 Babesia, 5,741 Toxoplasma, 6,981 erythrocytes	ResNet50 + GCN	Acc = 91.8, AUC = 91.8, Spec = 97.3
Proposed study	**Automated recognition of parasitic types in microscopic images**	**Parasite dataset (34,298 microscopic images)**	**Vision Transformer (ViT) Architecture**	**Acc = 99.70, Prec = 99.46, Rec = 99.40, Spec = 99.96, F1 = 99.45**



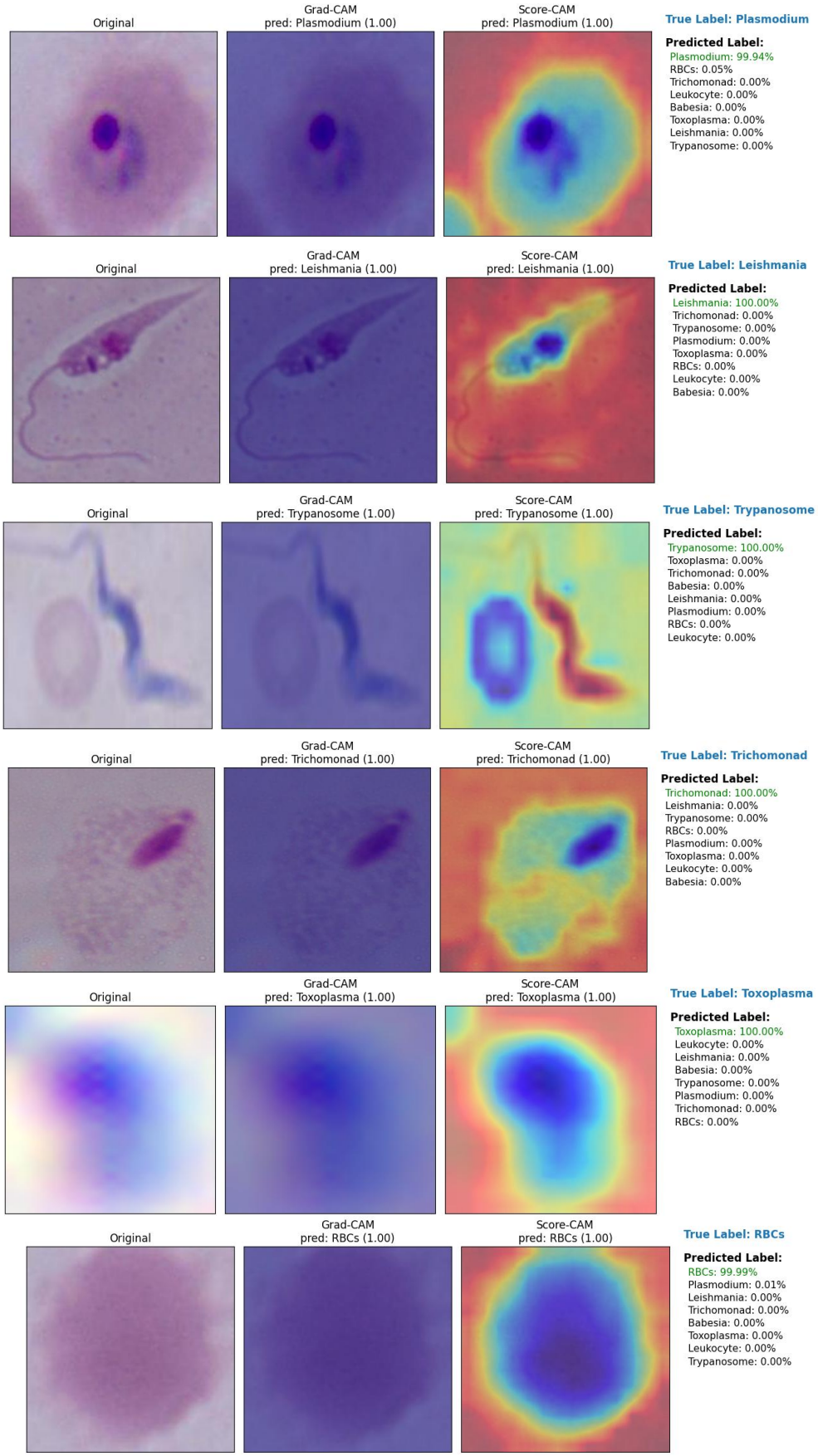


Fig. 8. Grad-CAM visualizations for representative samples

5. Conclusion

This study proposed a ViT model for the automated recognition of multiple parasitic types and host cells in microscopic images. Through the utilization of the global attention mechanism embedded within transformer architectures, the approach was able to capture subtle morphological variations that are often challenging for conventional CNN-based methods. The evaluation demonstrated consistently high performance across eight classes (Babesia, Leishmania, Leukocyte, Plasmodium, RBCs, Toxoplasma, Trichomonad, and Trypanosome). The proposed model achieved remarkable outcomes, achieving an overall accuracy of 99.70%, a precision of 99.46%, a recall of 99.40%, a specificity of 99.60%, and an F1-score of 99.45%. Confusion matrix analysis confirmed reliable class-wise predictions with minimal misclassification, while ROC-AUC values close to 1.0 across most classes underscored the model's excellent discriminative ability. Beyond numerical performance, interpretability was addressed through Grad-CAM and Score-CAM visualizations, which revealed that the ViT model consistently concentrated on biologically relevant areas, including parasite bodies, flagella, and nuclear components. Score-CAM, in particular, provided sharper localization compared to Grad-CAM, reinforcing trust in the decision-making mechanism of the model. Comparative analysis with state-of-the-art studies further highlighted that the proposed approach outperformed previous CNN, ensemble learning, and semi-supervised methods, establishing ViT as a superior framework for parasite recognition. In conclusion, the findings of this research demonstrate that Vision Transformers attain cutting-edge accuracy in automated parasite classification but also offer interpretability and robustness that are critical for clinical applications. This study advances the progress of computer-aided diagnostic systems by showing the potential of ViT-based architectures to support laboratory workflows, reduce diagnostic errors, and improve early detection of parasitic infections. Future research directions may involve enlarging the dataset with a wider range of staining approaches, enhancing localization for subtle classes like Plasmodium, and investigating lightweight ViT versions to facilitate adoption in healthcare environments with limited resources.

Supplementary Materials: The following supporting information can be downloaded at: https://www.kaggle.com/datasets/ahmedxc4/parasite-dataset/data?select=parasite-dataset

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding: This research was funded by the Institut Teknologi Sepuluh Nopember, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2025.

Acknowledgment: The authors would like to acknowledge the Department of Medical Technology, Institut Teknologi Sepuluh Nopember, for the facilities and support in this research. The authors also gratefully acknowledge financial support from the Institut Teknologi Sepuluh Nopember for this work, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2025.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Y. Ge *et al*., "The Global Prevalence of and Factors Associated with Parasitic Coinfection in People Living with Viruses: A Systematic Review and Meta-Analysis," *Pathogens*, vol. 14, no. 6, pp. 534-534, 2025, https://doi.org/10.3390/pathogens14060534.

[2] J. Zhang, Y. Sun, and J. Zheng, "The State of Art of Extracellular Traps in Protozoan Infections (Review)," *Frontiers in Immunology*, vol. 12, 2021, https://doi.org/10.3389/fimmu.2021.770246.

[3] W. Cheng, J. Liu, C. Wang, R. Jiang, M. Jiang, and F. Kong, "Application of image recognition technology in pathological diagnosis of blood smears," *Clinical and Experimental Medicine*, vol. 24, no. 181, 2024, https://doi.org/10.1007/s10238-024-01379-z.

[4] B. H. Gulumbe, A. Abdulrahim, S. K. Ahmad, K. A. Lawan, and M. B. Danlami, "WHO report signals tuberculosis resurgence: Addressing systemic failures and revamping control strategies," *Decoding Infection and Transmission*, vol. 3, p. 100044, 2025, https://doi.org/10.1016/j.dcit.2025.100044.

[5] T. Xu, N. T.-Umpon, and S. Auephanwiriyakul, "Staining-Independent Malaria Parasite Detection and Life Stage Classification in Blood Smear Images," *Applied Sciences*, vol. 14, no. 18, pp. 8402-8402, 2024, https://doi.org/10.3390/app14188402.

[6] A.-A. A.-Ramírez *et al*., "Malaria Cell Image Classification Using Compact Deep Learning Architectures on Jetson TX2," *Technologies*, vol. 12, no. 12, pp. 247-247, 2024, https://doi.org/10.3390/technologies12120247.

[7] E. Ahishakiye, F. Kanobe, D. Taremwa, B. A. Nantongo, L. Nkalubo, and S. Ahimbisibwe, "Enhancing malaria detection and classification using convolutional neural networks-vision transformer architecture," *Deleted Journal*, vol. 7, no. 612, 2025, https://doi.org/10.1007/s42452-025-06704-z.

[8] O. Elharrouss *et al*., "ViTs as backbones: Leveraging vision transformers for feature extraction," *Information Fusion*, vol. 118, p. 102951, 2025, https://doi.org/10.1016/j.inffus.2025.102951.

[9] S. Kumar, T. Arif, A. S. Alotaibi, M. B. Malik, and J. Manhas, "Advances Towards Automatic Detection and Classification of Parasites Microscopic Images Using Deep Convolutional Neural Network: Methods, Models and Research Directions," *Archives of Computational Methods in Engineering*, vol. 30, pp. 2013-2039, 2022, https://doi.org/10.1007/s11831-022-09858-w.

[10] R. C. Thomas *et al*., "Assessing rates of parasite coinfection and spatiotemporal strain variation via metabarcoding: Insights for the conservation of European turtle doves *Streptopelia* turtur," *Molecular Ecology*, vol. 31, no. 9, pp. 2730-2751, 2022, https://doi.org/10.1111/mec.16421.

[11] W. Xu, Q. Zhai, J. Liu, X. Xu, and J. Hua, "A lightweight deep-learning model for parasite egg detection in microscopy images," *Parasites & Vectors*, vol. 17, no. 454, 2024, https://doi.org/10.1186/s13071-024-06503-2.

[12] A. de S. Oliveira, M. G. F. Costa, M. das G. V. Barbosa, and C. F. F. C. Filho, "A new approach for malaria diagnosis in thick blood smear images," *Biomedical Signal Processing and Control*, vol. 78, p. 103931, 2022, https://doi.org/10.1016/j.bspc.2022.103931.

[13] M. Bhuiyan and M. S. Islam, "A new ensemble learning approach to detect malaria from microscopic red blood cell images," *Sensors International*, vol. 4, p. 100209, 2023, https://doi.org/10.1016/j.sintl.2022.100209.

[14] R. U. Khan, S. Almakdi, M. Alshehri, A. U. Haq, A. Ullah, and R. Kumar, "An intelligent neural network model to detect red blood cells for various blood structure classification in microscopic medical images," *Heliyon*, vol. 10, no. 4, p. e26149, 2024, https://doi.org/10.1016/j.heliyon.2024.e26149.

[15] S. Boit and R. Patil, "An Efficient Deep Learning Approach for Malaria Parasite Detection in Microscopic Images," *Diagnostics*, vol. 14, no. 23, p. 2738, 2024, https://doi.org/10.3390/diagnostics14232738.

[16] I. M. L. Chaharou, I. Lawani, T. Dagba, J. Degila, and H. A. Boubacar, "Image cropping for malaria parasite detection on heterogeneous data," *Journal of Microbiological Methods*, vol. 225, p. 107022, 2024, https://doi.org/10.1016/j.mimet.2024.107022.

[17] A. Alassaf and M. Y. Sikkandar, "Intelligent Deep Transfer Learning Based Malaria Parasite Detection and Classification Model Using Biomedical Image," *Computers, Materials & Continua*, vol. 72, no. 3, pp. 5273-5285, 2022, https://doi.org/10.32604/cmc.2022.025577.

[18] T. K. Kundu, D. K. Anguraj, and S. V. Sudha, "Modeling a Novel Hyper-Parameter Tuned Deep Learning Enabled Malaria Parasite Detection and Classification," *Computers, Materials & Continua*, vol. 77, no. 3, pp. 3289-3304, 2023, https://doi.org/10.32604/cmc.2023.039515.

[19] F. A. Muhammad, R. Sudirman, N. A. Zakaria, and S. N. S. S. Daud, "Morphology classification of malaria infected red blood cells using deep learning techniques," *Biomedical Signal Processing and Control*, vol. 99, p. 106869, 2025, https://doi.org/10.1016/j.bspc.2024.106869.

[20] Y. Ha, X. Meng, Z. Du, J. Tian, and Y. Yuan, "Semi-supervised graph learning framework for apicomplexan parasite classification," *Biomedical Signal Processing and Control*, vol. 81, pp. 104502, 2022, https://doi.org/10.1016/j.bspc.2022.104502.

[21] S. Li and Y. Zhang, "Microscopic Images of Parasites Species," *Mendeley Data*, vol. 3, 2020, https://doi.org/10.17632/38jtn4nzs6.3.

[22] L. Zedda, A. Loddo, and C. D. Ruberto, "A deep architecture based on attention mechanisms for effective end-to-end detection of early and mature malaria parasites," *Biomedical Signal Processing and Control*, vol. 94, p. 106289, 2024, https://doi.org/10.1016/j.bspc.2024.106289.

[23] J. D. K. H, "Implementation and Efficient Analysis of Preprocessing Techniques in Deep Learning for Image Classification," *Current Medical Imaging*, vol. 20, 2024, https://doi.org/10.2174/1573405620666230829150157.

[24] K. Maharana, S. Mondal, and B. Nemade, "A Review: Data Pre-Processing and Data Augmentation Techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91-99, 2022, https://doi.org/10.1016/j.gltp.2022.04.020.

[25] M. Krichen, "Convolutional Neural Networks: A Survey," *Computers*, vol. 12, no. 8, p. 151, 2023, https://doi.org/10.3390/computers12080151.

[26] K. Venkatesan *et al*., "Comparative analysis of resource-efficient YOLO models for rapid and accurate recognition of intestinal parasitic eggs in stool microscopy," *Intelligence-Based Medicine*, vol. 11, p. 100212, 2025, https://doi.org/10.1016/j.ibmed.2025.100212.

[27] R. Yang, R. Wang, Y. Deng, X. Jia, and H. Zhang, "Rethinking the Random Cropping Data Augmentation Method Used in the Training of CNN-Based SAR Image Ship Detector," *Remote Sensing*, vol. 13, no. 1, p. 34, 2020, https://doi.org/10.3390/rs13010034.

[28] D. O. Oyewola, E. G. Dada, S. Misra, and R. Damaševičius, "A Novel Data Augmentation Convolutional Neural Network for Detecting Malaria Parasite in Blood Smear Images," *Applied Artificial Intelligence*, vol. 36, no. 1, pp. 1-22, 2022, https://doi.org/10.1080/08839514.2022.2033473.

[29] A. Roddan, T. Czempiel, D. S. Elson, and S. Giannarou, "Calibration-Jitter: Augmentation of hyperspectral data for improved surgical scene segmentation," *Healthcare Technology Letters*, vol. 11, no. 6, p. 354, 2024, https://doi.org/10.1049/htl2.12102.

[30] K. Alomar, H. I. Aysel, and X. Cai, "Data Augmentation in Classification and Segmentation: A Survey and New Strategies," *Journal of Imaging*, vol. 9, no. 2, p. 46, 2023, https://doi.org/10.3390/jimaging9020046.

[31] M. Jeong, M. Yang, and J. Jeong, "Hybrid-DC: A Hybrid Framework Using ResNet-50 and Vision Transformer for Steel Surface Defect Classification in the Rolling Process," *Electronics*, vol. 13, no. 22, p. 4467, 2024, https://doi.org/10.3390/electronics13224467.

[32] M. Trigka and E. Dritsas, "A Comprehensive Survey of Deep Learning Approaches in Image Processing," *Sensors*, vol. 25, no. 2, p. 531, 2025, https://doi.org/10.3390/s25020531.

[33] Y. Al Khalil, S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images," *Medical Image Analysis*, vol. 84, p. 102688, 2022, https://doi.org/10.1016/j.media.2022.102688.

[34] Y. Wang, Y. Deng, Y. Zheng, P. Chattopadhyay, and L. Wang, "Vision Transformers for Image Classification: A Comparative Survey," *Technologies*, vol. 13, no. 1, p. 32, 2025, https://doi.org/10.3390/technologies13010032.

[35] O. Chibuike and X. Yang, "Convolutional Neural Network-Vision Transformer Architecture with Gated Control Mechanism and Multi-Scale Fusion for Enhanced Pulmonary Disease Classification," *Diagnostics*, vol. 14, no. 24, p. 2790, 2024, https://doi.org/10.3390/diagnostics14242790.

[36] M. D. Niz, S. S. Pereira, D. Kirchenbuechler, L. Lemgruber, and C. Arvanitis, "Artificial intelligence-powered microscopy: Transforming the landscape of parasitology," *Journal of Microscopy*, 2025, https://doi.org/10.1111/jmi.13433.

[37] O. Katar and O. Yildirim, "An explainable Vision Transformer model based white blood cells classification and localization," *Diagnostics*, vol. 13, no. 14, p. 2459, 2023, https://doi.org/10.3390/diagnostics13142459.

[38] Y. S. Ju, Z. W. Geem, and J. S. Lim, "Attention Score Enhancement Model Through Pairwise Image Comparison," *Applied Sciences*, vol. 14, no. 21, p. 9928, 2024, https://doi.org/10.3390/app14219928.

[39] M. Ennab and H. Mcheick, "Advancing AI Interpretability in Medical Imaging: A Comparative Analysis of Pixel-Level Interpretability and Grad-CAM Models," *Machine Learning and Knowledge Extraction*, vol. 7, no. 1, p. 12, 2025, https://doi.org/10.3390/make7010012.

[40] Q. Mastoi *et al*., "Explainable AI in medical imaging: an interpretable and collaborative federated learning model for brain tumor classification," *Frontiers in Oncology*, vol. 15, 2025, https://doi.org/10.3389/fonc.2025.1535478.

[41] A. Jierula, S. Wang, T.-M. OH, and P. Wang, "Study on Accuracy Metrics for Evaluating the Predictions of Damage Locations in Deep Piles Using Artificial Neural Networks with Acoustic Emission Data," *Applied Sciences*, vol. 11, no. 5, p. 2314, 2021, https://doi.org/10.3390/app11052314.

[42] S. A. Hicks, I. Strümke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, and S. Parasa, "On evaluation metrics for medical applications of artificial intelligence," *Scientific Reports*, vol. 12, no. 1, p. 5979, 2022, https://doi.org/10.1038/s41598-022-09954-8.

[43] T. F. Monaghan *et al*., "Foundational Statistical Principles in Medical Research: Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value," *Medicina*, vol. 57, no. 5, p. 503, 2021, https://doi.org/10.3390/medicina57050503.

[44] O. A. M.-López, N. Kismiantini, A. Alemu, A. M.-López, J. C. M.-López, and J. Crossa, "Balancing Sensitivity and Specificity Enhances Top and Bottom Ranking in Genomic Prediction of Cultivars," *Plants*, vol. 14, no. 3, p. 308, 2025, https://doi.org/10.3390/plants14030308.

[45] M. C. H. Lee, J. Braet, and J. Springael, "Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted F1-Scores," *Applied Sciences*, vol. 14, no. 21, p. 9863, 2024, https://doi.org/10.3390/app14219863.

[46] C. B. Delahunt, N. Gachuhi, and M. P. Horning, "Metrics to guide development of machine learning algorithms for malaria diagnosis," *Frontiers in Malaria*, vol. 2, 2024, https://doi.org/10.3389/fmala.2024.1250220.

[47] G. P. Reddy, D. Rohan, S. M. A. Kareem, Y. V. P. Kumar, K. P. Prakash, and M. Janapati, "A Custom Convolutional Neural Network Model-Based Bioimaging Technique for Enhanced Accuracy of Alzheimer's Disease Detection," *Engineering Proceedings*, vol. 87, no. 1, p. 47, 2025, https://doi.org/10.3390/engproc2025087047.

[48] M. Z. Naser, "From failure to fusion: A survey on learning from bad machine learning models," *Information Fusion*, vol. 120, p. 103122, 2025, https://doi.org/10.1016/j.inffus.2025.103122.

[49] M. R. Ahmed *et al*., "Hierarchical Swin Transformer Ensemble with Explainable AI for Robust and Decentralized Breast Cancer Diagnosis," *Bioengineering*, vol. 12, no. 6, p. 651, 2025, https://doi.org/10.3390/bioengineering12060651.

[50] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies*, vol. 12, no. 7, p. 113, 2024, https://doi.org/10.3390/technologies12070113.

[51] G. Madhu, A. W. Mohamed, S. Kautish, M. A. Shah, and I. Ali, "Intelligent diagnostic model for malaria parasite detection and classification using imperative inception-based capsule neural networks," *Scientific Reports*, vol. 13, no. 13377, 2023, https://doi.org/10.1038/s41598-023-40317-z.

[52] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," *Frontiers in Public Health*, vol. 11, no. 1273253, 2023, https://doi.org/10.3389/fpubh.2023.1273253.

[53] S. Mahmood, R. Hasan, S. Hussain, and R. Adhikari, "An Interpretable and Generalizable Machine Learning Model for Predicting Asthma Outcomes: Integrating AutoML and Explainable AI Techniques," *World*, vol. 6, no. 1, p. 15, 2025, https://doi.org/10.3390/world6010015.

[54] W. Fang *et al*., "Diagnosis of invasive fungal infections: challenges and recent developments," *Journal of biomedical science*, vol. 30, no. 42, 2023, https://doi.org/10.1186/s12929-023-00926-2.

[55] R. A. Taylor *et al*., "Leveraging artificial intelligence to reduce diagnostic errors in emergency medicine: Challenges, opportunities, and future directions," *Academic Emergency Medicine*, vol. 32, no. 3, pp. 327-339, 2024, https://doi.org/10.1111/acem.15066.

[56] C. M. Tsai and J.-D. Lee, "Dynamic Ensemble Learning with Gradient-Weighted Class Activation Mapping for Enhanced Gastrointestinal Disease Classification," *Electronics*, vol. 14, no. 2, p. 305, 2025, https://doi.org/10.3390/electronics14020305.

[57] Y. Zou and P. Miao, "Explainable AI-enabled hybrid deep learning architecture for breast cancer detection," *Frontiers in Immunology*, vol. 16, 2025, https://doi.org/10.3389/fimmu.2025.1658741.

[58] E. H. Houssein, A. M. Gamal, Eman, and E. Mohamed, "Explainable artificial intelligence for medical imaging systems using deep learning: a comprehensive review," *Cluster Computing*, vol. 28, no. 469, 2025, https://doi.org/10.1007/s10586-025-05281-5.

[59] R. Fernandez, A. G.-Ponce, R. F.-Beltran, and G. G.-Mateos, "Symmetry in Explainable AI: A Morphometric Deep Learning Analysis for Skin Lesion Classification," *Symmetry*, vol. 17, no. 8, p. 1264, 2025, https://doi.org/10.3390/sym17081264.

[60] Z. Li and O. Dib, "Empowering Brain Tumor Diagnosis through Explainable Deep Learning," *Machine Learning and Knowledge Extraction*, vol. 6, no. 4, pp. 2248-2281, 2024, https://doi.org/10.3390/make6040111.

[61] A. B.-Montero *et al*., "Artificial intelligence and machine learning for medical imaging: A technology review," *Physica Medica*, vol. 83, pp. 242-256, 2021, https://doi.org/10.1016/j.ejmp.2021.04.016.