Deep Learning-Driven Scalable and High-Precision Malaria Detection from Microscopic Blood Smear Images

Dr. N. Kannaiya Raja[1], Divya Rohatgi[2], Venkata Lalitha Narla[3], Ganesh Kumar Anbazhagan[4],
R. Aroul Canessane[5], Drakshayani Sriramsetti[6], Prof. Ts. Dr. Yousef A.Baker El-Ebiary[7]
Sr. Associate Professor, School of Computing Science and Engineering, VIT Bhopal University, Bhopal, Madhya Pradesh, India[1]
Associate Professor, Bharati Vidyapeeth Deemed to be University,
Department of Engineering and Technology, Maharashtra, India[2]
Associate Professor, Department of Electronics and Communication Engineering, Aditya University, Surampalem, AP, India[3]
Department of Microbiology, Saveetha Medical College and Hospital,
Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai – 602 105, Tamil Nadu, India[4]
Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology,
Chennai, Tamil Nadu, India[5]
Assistant Professor, Dept. of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur Dist., Andhra Pradesh - 522502, India[6]
Faculty of Informatics and Computing, UniSZA University, Malaysia[7]

***Abstract*—Malaria continues to be a life-threatening disease, especially in tropical and low-resource regions, where timely and accurate diagnosis remains a major challenge. Traditional diagnostic approaches like manual microscopy are not only time-consuming and expertise-dependent but also prone to subjective errors. Existing deep learning methods, such as Convolutional Neural Networks (CNNs), ResNet, and Vision Transformers (ViT), struggle to generalize across variations in staining, resolution, and morphology, leading to misclassification and reduced diagnostic reliability. To overcome these limitations, this study proposes a novel hybrid architecture, Swin-Siamese, which integrates the hierarchical self-attention mechanism of the Swin Transformer with the contrastive similarity learning capability of the Siamese Neural Network. This unique combination enables the model to capture both global and local spatial patterns while accurately distinguishing infected from uninfected blood smear images. The model is implemented using TensorFlow and PyTorch, and trained on a publicly available malaria dataset comprising 13,152 training, 626 validation, and 1,253 test images. Experimental results demonstrate a 3.1% improvement in accuracy over traditional CNNs, achieving 95.3% accuracy, 95.1% precision, 95.4% recall, 95.2% F1-score, and an AUC-ROC of 0.97. This significant performance gain highlights the model's scalability, interpretability, and real-time applicability in clinical and field-deployable diagnostic systems, offering a powerful solution for malaria screening in underserved regions.**

Keywords*—Automated diagnosis; blood smear images; contrastive learning; deep learning; malaria detection

I. Introduction

Malaria continues to pose a major danger to health in many tropical and subtropical regions [1]. In 2021, the WHO estimated that 247 million cases of malaria and 619,000 deaths occurred, and sub-Saharan Africa experienced the biggest incidence. Malaria causes significant problems for children younger than five years. Microscopy, rapid diagnostic tests (RDTs) and polymerase chain reaction (PCR) each have certain drawbacks [2], [3]. Many RDTs offer limited sensitivity, and PCR activities cannot be carried out without advanced laboratory resources. Direct look at Plasmodium with a microscope under a blood smear is still the main way of diagnosing malaria [4], [5]. Unfortunately, this method is a time-consuming process and is slow, while being subject to a number of human errors [6]. Although PCR-based methods are highly accurate, they are laboratory-suited and require sophisticated equipment. Considering these challenges, the availability of automatically, AI-driven diagnostic systems capable of automatically analyzing infected blood smear images and classifying them efficiently is of great importance [7]. Malaria detection using CNN-based models has shown strong classification performance, which allows for fast and low human intervention in the process of malaria diagnosis [8]. Furthermore, CNN-based models may have difficulties in differentiation with blood smear images, variations in staining, resolution and lighting conditions [9]. In particular, these limitations indicate the need for more advanced architectures of deep learning (DL) that can make better feature representation and make better classification accuracy [10]. Originally, transformer-based models have been created for NLP, and recently adapted for computer vision problems, giving rise to ViTs [11]. The Swin Transformer is one of the most advanced transformer-based architectures for vision tasks and uses hierarchical feature extraction along with shifted window attention to enhance computational efficiency as well as performance [12]. Swin Transformers are different from traditional ViTs since they only apply self-attention to the whole image, rather than the whole image at once. The image is divided into non-overlapping windows to extract local features and at the same time such windows can interact with their neighbor windows. In this hierarchical way, Swin Transformers are more

efficient and scalable in handling high resolution medical images. Another powerful architecture in DL is the Siamese Network, which has been applied in similarity learning and classification tasks [13]. Siamese Networks are different from traditional classification networks, which operate independently on images. By combining a Swin Transformer with a Siamese Network, it adds two key advantages: the unparalleled feature extraction and contrastive learning capability of the ensemble Crohn and contrastive learning to maximum classification performance in malaria detection [14]. In this hybrid model, the Swin Transformer is used as the feature extractor to extract the rich representations of blood smear images, and is enhanced by the Siamese Network for fine classification accuracy on infected and uninfected samples, respectively [15]. This study's research objective is to build an automated malaria detection system using Swin Transformer and Siamese Network architectures so that classification accuracy and robustness can be enhanced [16] [17]. The key goals are to design a Swin Transformer-based malaria detection model, merge a Siamese Network to practice contrastive study and maximize the model for actual world implementation in low-resource settings. This study combines the advantages of the self-attention mechanisms and contrastive learning to build a highly accurate, scalable and efficient malaria diagnosis system which is deployable in the real-world clinical settings, especially in regions devoid of expert microscopists.

A. Research Significance

Malaria is an important health problem across the world, particularly in countries with low or limited healthcare staff and equipment. While conventional methods such as microscopy and PCR are fine when resources are available, they prove either slow or unnatural for settings with limited resources. Applying AI and DL to diagnoses could quickly, accurately and affordably improve malaria diagnosis. It matters because this research introduces and makes use of Swin Transformer and Siamese Neural Network to help diagnose malaria parasites through blood smear images. I use a Swin Transformer model that includes a hierarchical self-attention mechanism together with Siamese Networks that rely on comparison and similarity. This hybrid model overcomes various weaknesses of general CNN-based models, including poor feature generalization and inefficiency in dealing with image variability. The high-performance metrics of the proposed model—95.3% accuracy and 0.97 AUC-ROC—prove its validity in real-world clinical and remote applications. Therefore, this study adds to the literature in medical diagnostics by suggesting a resilient, scalable, and effective AI-driven solution that improves early treatment and detection of malaria, hence complementing global disease control initiatives.

B. Research Motivation

The main drive behind this work arises from the necessity to develop malaria diagnosis in areas where conventional healthcare infrastructure is under-equipped. If preventable and treatable, malaria still kills hundreds of thousands of people every year, mainly because diagnosis is slow and unreliable. Manual microscopic examination, even if standard in most practices, is time and expert-consuming—factors not always found in rural or developing communities. Current machine learning approaches, especially CNN-based approaches, have been promising but, nonetheless, had their shortcomings like inadequate feature extraction in noisy or complicated images and very high reliance on large amounts of labeled data. With these shortfalls in consideration, this research is spurred to test how much potential there is for transformer-based models, specifically the Swin Transformer, whose forte has been its capacity to learn very fine-grained spatial information through hierarchical attention mechanisms. Furthermore, incorporating a Siamese Neural Network adds contrastive learning, which improves the model's ability to differentiate between infected and uninfected samples even for changing imaging conditions. The motivation is further supported by the possibility of creating an instrument that not only has high accuracy but is also computationally efficient—thus fit for deployment in low-resource environments. Finally, the aim is to move AI-based malaria diagnosis closer to actual, field-level application.

C. Problem Statement

Malaria continues to pose a severe health risk in low-resource areas, for which timely and precise diagnosis must decrease mortality. All the current models, like Swin Transformer with Siamese Networks, are still hampered in the capacity of generalization across a wide range of microscopy environments and geography [18]. The models tend to perform poorly under inadequately addressing biases due to differences in staining methods, imaging parameters, and patient populations. Also, although Transformer-based models provide better feature extraction, they are computationally intensive, which limits their realistic usage in clinic settings with limited resources [19]. The majority of existing models also cater only to binary classification and do not have the ability to differentiate among various Plasmodium species, something important for proper treatment. Additionally, the interpretability of DL models remains a concern, since their black-box nature does not allow them to gain clinical trust and acceptance. Moreover, there is also a reliance on high-quality annotated datasets, which are scarce in endemic areas, so robustness and scalability are pretty difficult to achieve [20]. These shortcomings limit the integration of AI-assisted malaria diagnostic systems in healthcare settings. The current study proposes to address these gaps by developing an augmented data augmentation, contrastive learning hybrid Swin Transformer–Siamese model to improve the diagnosis accuracy, generalizability, explainability, and computational expense for real-life use cases.

D. Key Contribution

- This study introduces a new hybrid framework by integrating the Swin Transformer and Siamese Network to improve automated malaria diagnosis from microscopic blood smear images.
- Contrastive learning is embedded through a Siamese approach to effectively identify infected and uninfected cells, even in difficult imaging conditions.
- The model is engineered to enhance generalization and robustness to variations in staining, resolution, and lighting, resolving major limitations of current CNN-based methods.
- By emphasizing computational efficiency and scalability, the method being proposed is optimized for

implementation within real-world, resource-limited clinical environments.

The Swin-Siamese model has definite benefits over traditional methods. In contrast to CNN and ResNet models that mainly extract local features, the hierarchical attention mechanism of Swin Transformer allows for efficient extraction of global and local patterns to enhance robustness under different staining and imaging conditions. The inter-class separability is considerably boosted by the contrastive learning component of the Siamese Network, resulting in fewer false negatives and increased diagnostic reliability, compared with baseline models. In addition, the model's moderate computational profile and rapid inference time ensure efficient deployment in low-resource clinical settings, filling the space between cutting-edge AI methods and real-world medical diagnostics.

E. Rest of the Section

Section II is the review of the existing research relevant to the DL technique applied to malaria detection and examines the limitations of those works. Section III describes the SwinSiaNet framework, which includes preprocessing, the Swin transformer, the Siamese Network, and the loss details. Section IV will show the experimental results, comparative evaluations, ablation studies, and feasibility study for deployment. Lastly, Section V concludes this study and describes future research goals to improve real-time resource-aware malaria diagnosis.

II. Related Work

First, a DL model is built to sort through and recognize different malaria parasite types accurately from both thin and thick peripheral blood smear microscopic images. The other question is to determine which specimen has a better chance of identifying parasites in the peripheral blood smears. Following this, the study assesses the efficacy of the approach relative to well-known transfer learning models. Therefore, a convolutional neural network is suggested for highly accurate malaria parasite prediction from thick and thin peripheral blood smear images seen under a microscope. The measurements for model performance with the standard evaluations were good. Improvement in the model was observed with 96.97% accuracy, 97.00% precision and 97.00% sensitivity when thick peripheral smears were used. With the right peripheral blood smear, faster, more precise smear preparation, and patient diagnosis can be done in malaria-prone regions [21]. However, because of the variability in staining and image quality among laboratories, the model may not perform perfectly in each environment. The model may depend on the information provided by a dataset so much that it is not widely applicable in practical medicine.

Looking at malaria parasitemia helps doctors decide on the level of disease severity and plans the most appropriate therapy. For a long time, malaria parasitemia has been identified through thick smear blood film microscopy. It may accurately measure parasite numbers faster than any other method yet, but it has been declared dissatisfactory due to being laborious, requiring a high degree of expertise and taking a lot of time to complete. Low-funded technical staff and high levels of endemicity are both major obstacles in many developing countries. Yet, this research provides a solution by using an approach that locally identifies and calculates both WBCs and parasites present. The approach was to set up computer vision models by training on thick blood smear images that had been annotated. The pre-trained DL models, Faster R-CNN and SSD, have been applied to build computer vision models that use acquired digital images. Not having enough data led to the use of augmentation to strengthen the model's performance. They have found that the method is able to correctly estimate both the number of parasites and WBCs with strong accuracy and recall. The findings agreed well with the counts detected by the observers using the microscopy. It is possible to use this approach in devices where there are few Microscopy Experts and a large number of patients needing diagnosis. The approach I have proposed works best when annotated datasets are available, but those might not exist all the time [22]. However, deploying such systems in mobile devices might require them to work faster while using less energy.

To handle malaria, one needs to test and diagnose it quickly and estimate the parasite load. Microscopic examination of a PB smear forms the best approach for diagnosing malaria. Even so, this procedure takes a long time. That's why an automated system is set up using the microscope to measure and spot malaria parasites. The system uses a microscope, a plastic chip, fluorescent dye and an image analysis programmer. Results for linearity, precision and limit of detection of my analysis were compared to those from traditional microscopic PB smear tests and flow cytometry. The system showed satisfactory linearity by demonstrating similar results with the Plasmodium falciparum culture and the Plasmodium vivax-infected sample. This analysis showed that the %CV at every parasitemia level was precise, and for all parasitic loads, the %CV in the system was lower than in microscopic examination. Assessment of the limit of detection pointed out that the likelihood of detecting the parasite was 0.00066112% and a good match was seen among all the techniques. The system was both highly specific and highly sensitive, correctly identifying every P. vivax and P. falciparum sample [23]. Several characteristics of the automated malaria parasite detection system help it detect parasites more quickly and monitor their density better than manual microscopy. Even so, the process depends on using fluorescence dyes, and you will likely need to take extra preparation steps and have some specialized reagents.

To support malaria diagnosis of thick smear microscope pictures, Plasmodium VF-Net is built. The method which is employed can identify if a person has malaria and which type of the disease is present: Plasmodium falciparum or Plasmodium vivax. Extracted candidates from Plasmodium parasites are initially found as regions with Mask RCNN, then the candidates go through a ResNet50 classifier and a new species detection approach uses the number of patches discovered and the probability from all the patient images. It is a tough job to describe a patient level decision: the parasites are too small, many species look the same, there are many types of color and lighting and not all samples are stained uniformly. More specifically, a dataset of 350 patients is used, containing over 18 million images and publicly release the images used in this manuscript alongside this publication. This system achieves accuracy greater than 90% at both the image and patient levels.

Excellent performance from the Plasmodium-Net requires using strong annotated datasets; these are not always available in complex and diverse practical situations [24]. Furthermore, how stains are performed and the quality of images produced by different laboratories can affect how well the model generalizes and remains robust.

Every single day, lots of people across the globe are affected by malaria. To diagnose malaria, a doctor routinely looks at patient's blood under a microscope to see if the malaria parasite is present. It is generally a slow and faulty process. As a result, malaria type and its progression stages can now be detected and correctly sorted out. Here, YOLOv5 and Yolo v4, two detection models, were proposed to differentiate the stage of malaria and the type of parasite present. Two different datasets are chosen to ensure the task can be tested for both the stage and the type of parasite. The selected data consists of microscopic pictures of red blood cells, some infected and some healthy. The findings were grouped by whether the infection was caused by one kind of malarial parasite or another and what stage of malaria it represented. The data used was annotated by hand with the labeling tool. After that, the models were improved to make the image training more effective. Authors have found YOLOv5 and scaled YOLOv4 effective at classifying the type of parasite detected. The Scaled YOLOv4 and YOLOv5 had accuracies of 83% and 78.5% respectively. Maybe the proposed models can support medical professionals in both diagnosing malaria and anticipating what stage it is in. Yet, these models still depend on manually created datasets, which could introduce inconsistencies that reduce their performance. It also suggests that the detection performance can improve for identifying the stages that most closely resemble the parasite. The proposed models, however, still leaned on the use of manually annotated datasets, which may contribute to labeling inconsistencies that will affect the performance [25]. Furthermore, 83% accuracy for scaled YOLOv4 also implies there is room for improvement to identify the parasite stages of high similarity.

Despite years of efforts to eliminate the disease, malaria is still a major global health problem, especially in underdeveloped countries where poor access to health care is prevalent. DL and ML-based approaches for the detection and classification of malaria from blood smears become necessary as traditional microscopic examination is labor-intensive and needs expertise of an expert. A number of other proposed models for malaria parasite classification, detection, and estimation of parasitemia are transfer learning models, CNNs, and object detection frameworks. Specific models, for example, MobileNetV2 and DenseNet-201, achieved high precision rates of 97.06% and 99.40%, separately. There were also other approaches that other people used to increase detection performance such as combining multiple classifiers like Plasmodium VF-Net and ROENet, for example. In automated microscopy systems that combine fluorescent dye and image analysis, sensitivity and specificity were 100%. Although these advancements solve these challenges, there remains challenges such as the use of high-quality annotated datasets, variations in staining techniques and the computational difficulties in time for real-time implementation in restricted environments. It also outperformed ourselves on cross dataset experiment and highlighted the need for more robust and generalized model. While DL has promise in boosting malaria diagnosis, more work is required for clinical implementation to be widespread.

To date, malaria is still a major global health concern, especially in developing regions where a timely and accurate diagnosis is highly needed to enable prompt treatment. In the gold standard for malaria detection, pathologists examine blood smears to look at the details done by trained pathologists. Nevertheless, this method is resource and time-consuming, error-prone, and requires specific expertise in these remote areas. Despite its promise to automate malaria detection, conventional DL models like CNNs usually do not perform very well with respect to feature extraction, generalization, and handling of the variations in blood smear images. This study proposes a more advanced DL approach by coupling the Siamese Network with the powerful network design of Swin Transformer [26]. The hierarchical self-attention-based Swin Transformer makes a better contribution to the feature extraction, while the Siamese Network helps with the similarity learning in discriminating infected and uninfected cells [27]. The goal is to come up with a model surpassing the performance of the existing CNN-based models and one that is computationally viable for deployment in real-world healthcare settings.

Current research in malaria detection has investigated CNNs, object detection architectures such as YOLOv5 and Faster R-CNN, and combination approaches like Plasmodium VF-Net and MobileNetV2, with accuracies of up to 99.4%. Although these approaches exhibit high performance, they tend to fail during generalization across staining methods, image quality, and small annotated datasets. Most models are heavily dependent on thick or thin smears and handcrafted annotations, with reduced scalability and reliability. Certain models incorporate fluorescent dyes or run on mobile devices, but require preprocessing steps or intensive computing. In spite of these improvements, issues such as interpretability, computational cost, and flexibility in resource-limited environments remain. This makes a stronger, more scalable, and interpretable solution necessary, resulting in the suggested Swin-Siamese model for enhanced malaria detection accuracy and efficiency.

III. Research Methodology

The Swin Transformer and Siamese Network are combined for the malaria detection. The process starts with data collection and then preprocessing steps such as resizing, normalization, and augmentation are performed on the dataset. The hierarchical features are extracted by the Swin Transformer via shifted window attention. The learning of similarity and difference in the image pairs is done by the Siamese Network. The contrastive loss function is applied on feature embeddings and processed. The gradient based optimization with GPU acceleration is used on the model. This gives improved performance of the malaria detection and feature learning. This also offers a reliable DL-based diagnosis. The workflow of the proposed study is given in Fig. 1.

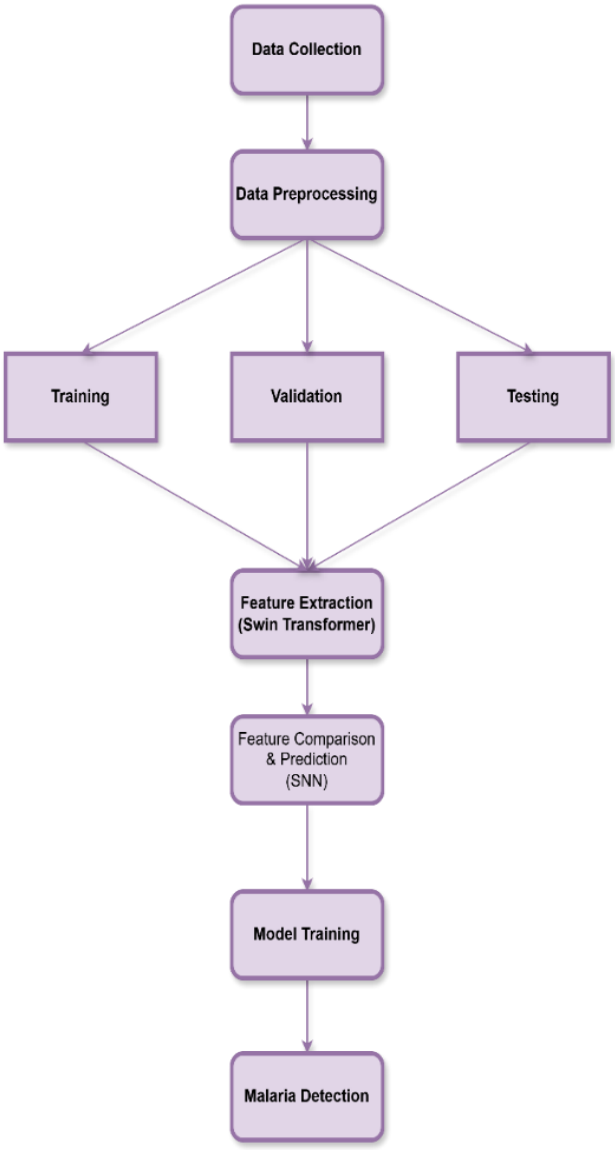

Fig. 1. Flow of malaria parasite detection.

A. Data Collection

The Malaria Detection dataset is collected from the Kaggle dataset [28]. The data used are of high-resolution microscopic blood smear images that are labelled malaria-infected or uninfected images. The DL model needs to be well-trained, validated, and tested on this data since it is necessary for the DL model to generalize well in real life. A properly curated dataset assists the model in learning the intricate patterns created by the malaria parasites and, therefore, the diagnosis is improved. Three subsets of the dataset are the training set, validation set and test set. It contains 13,152 images in the training set that are employed in training the model and depict the learned features. The validation set contains 626 images that will be utilized to fine-tune the hyperparameters and prevent overfitting during the training process by observing how good the model performs in training. The last test set contains 1,253 images, which serve as an independent estimate of the final model's generalization capability.

B. Data Pre-processing

To retain consistency and suitability for DL input, every image of the malaria data set is downscaled to 224×224 pixels. Downscaling was done after a min–max normalization, i.e. the image pixel intensities were rescaled to have a value between [0, 1], thereby stabilizing the gradient descent during training. This normalization is calculated as Eq. (1):

$$I_{norm}(x,y) = \frac{I(x,y)-I_{min}}{I_{max}-I_{min}} \quad (1)$$

where, $I(x,y)$ is the original pixel value, and I_{min}, I_{max} are the minimum and maximum intensity values, respectively.

Extensive data augmentation is employed to boost model generalizability and robustness. The augmented data is produced through random rotations ($\pm 15^0$), horizontal/vertical flips, contrast adjustments, and added Gaussian noise. They are labeled as the transformation composition $A(I)$, as in Eq. (2):

$$\tilde{I} = A(I) = T_{flip} \cdot T_{rotate} \cdot T_{contrast} \cdot T_{noise}(I) \quad (2)$$

It helps the model to learn distortion-invariant features in field-acquired microscopy images.

C. Feature Extraction via Swin Transformer

Hierarchical feature extraction and contrastive learning enable a very compact and efficient DL model for malaria detection by leveraging the Swin Transformer together with a Siamese Network. In contrast to local receptive fields typical of CNNs, Transformers use the self-attention mechanism to capture long-range dependencies; Swin Transformer further reduces the computation cost by utilizing the shifted windowing learning mode, which balances the global and the local feature learning. The architecture is of a patch embedding layer, self-attention layers, MLPs and normalization layers to extract a robust feature from the microscopic blood smear image. With a Siamese Network, this capability is further improved as it compares image pairs using the same branches with shared weights and measures their similarity with distance metrics such as the Euclidean distance, which results in more precision. The hyperparameters, which are key to performance, are tuned in order to optimize the performance, resulting in better generalization across many different image conditions. Transformers outperform CNNs in capturing fine-grained details and discarding false positives, as well as in generalizing to imaging variations. Being a hybrid of Swin Transformer and Siamese Network, the model can be used in medical diagnostics with high precision, reliability and adaptability and can become a promising way to detect malaria at an early stage. The architecture of the Swin transformer is given in Fig. 2.

The Swin Transformer is used to extract hierarchical features from blood smear images. The input image is split into non-overlapping patches, which are embedded and passed through layers of shifted window based self-attention, allowing for better exploitation of local and global features. The self-attention in a shifted window is computed as Eq. (3):

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}} + B\right)V \quad (3)$$

where, Q, K, and V are the 'query', 'key', and 'value' matrices, d_k is the key dimension, and B is the relative

positional encoding bias. Each transformer block uses multi-layer perceptrons (MLPs) to learn complex feature mappings, defined as Eq. (4):

$$\text{MLP}(x) = \sigma(W_2(ReLU(W_1 x + b_1)) + b_2 \quad (4)$$

where, W_1, W_2 are weight matrices, b_1, b_2 are biases, and σ denotes a non-linear activation function.

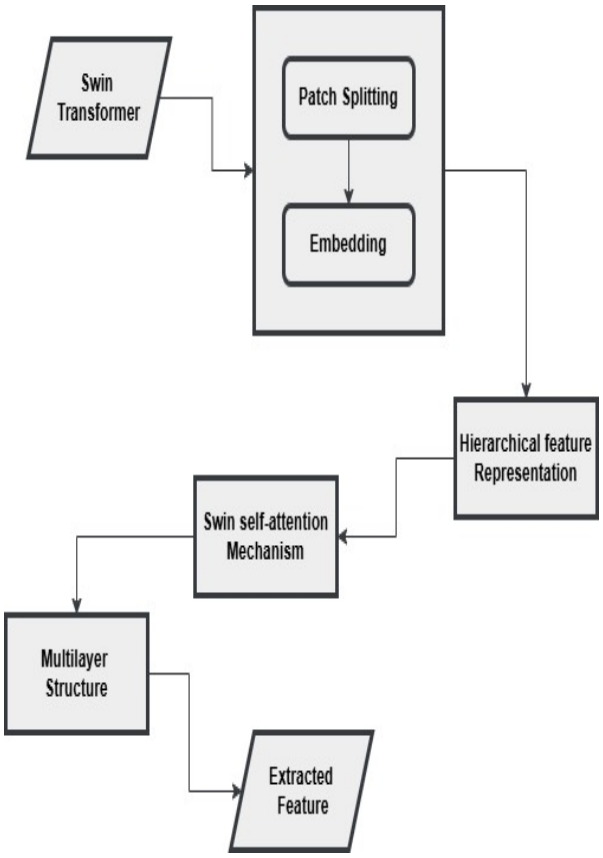

Fig. 2. Architecture of Swin transformer.

D. Similarity Learning via Siamese Network

The Siamese Network is a DL architecture specifically designed for similarity learning, where two identical neural networks process paired inputs to determine their similarity. Both networks share weights, ensuring consistent feature extraction and enabling the model to effectively compare images based on learned patterns. This architecture is widely used in image recognition, verification, and anomaly detection, making it particularly useful for medical imaging applications such as malaria detection. This characteristic improves the model's ability to handle imbalanced datasets, a common issue in medical diagnostics. Furthermore, the Siamese Network excels in scenarios with limited training data, as it focuses on learning discriminative relationships rather than memorizing specific features. By enhancing the robustness and accuracy of classification, this approach is highly suitable for real-world diagnostic applications, reducing dependency on large datasets while ensuring reliable malaria detection and other medical image-based diagnoses.

A Siamese Network with two identical Swin Transformer encoders, sharing weights, is used to improve classification performance in small or unbalanced datasets. Each encoder applies to one of the images and generates deep feature embeddings $f(x_1)$ and $fx_2)$ for the pair of images, x_1 and x_2 respectively. The Euclidean distance between embeddings will be computed as Eq. (5):

$$D(x_1, x_2) = \| f(x_1) - fx_2) \| 2 \quad (5)$$

For image-pair classification, cosine similarity is also evaluated to measure the directional closeness of features, as given in Eq. (6):

$$Sim(x_1, x_2) = \frac{x_1 . x_2}{\|x_1\| \|x_2\|} \quad (6)$$

E. Swin Transformer with Siamese Neural Network for Malaria Detection

Swin Transformer and Siamese Network together bring a powerful malaria detection solution by utilizing the strengths of both architectures. It helps the model distinguish little between infected and noninfected cells. The other way is to use the Siamese Network for similarity learning, which takes two images and compares the feature embeddings of two separate identical networks. The modifications enable the model to concentrate on learning what the distinguishing characteristics of uninfected and infected blood smear images are, thereby leading to higher accuracy even with less data. The use of shared weights is to make the feature extraction practice consistent in the Siamese Network, whereas the use of a contrastive loss function pulls similar samples closer and pushes the dissimilar samples apart to make the network robust. This method integrates Swin Transformer's efficient attention mechanism with Siamese Network's pairwise comparison ability to achieve very high malaria detection accuracy, with high scalability, efficiency and data requirement. The flow of the model Swin-Siamese is given in Fig. 3.

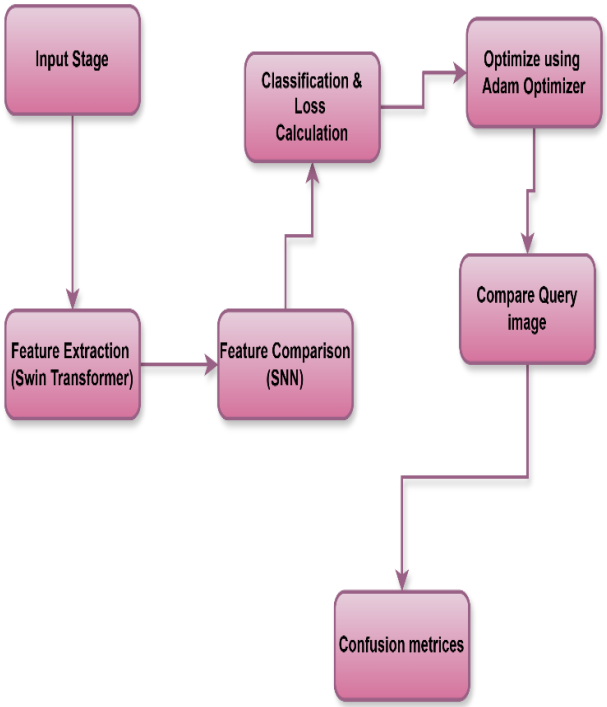

Fig. 3. Swin transformer with Siamese neural network for malaria detection.

F. Composite Loss Function and Optimization

To successfully train a model, the composite loss function, L_{total}, combines a cross-entropy loss for classification and, a contrastive loss for learning similarity. The contrastive loss function is defined as Eq. (7):

$$L_{contrastive} = (1-Y)\frac{1}{2}D^2 + (Y)\frac{1}{2}max(0, m-D)^2 \quad (7)$$

where, $Y = 0$ for similar pairs and $Y = 1$ for dissimilar ones, and mmm is a margin parameter. The final training objective aggregates multiple loss terms, as given in Eq. (8):

$$L_{total} = \propto L_{Classification} + \beta L_{contrastive} \quad (8)$$

where, α and β are weighting factors that balance the two losses.

G. Influence of Hyperparameters

The Swin-Siamese model's performance is sensitive to a number of critical hyperparameters. The learning rate controls the convergence rate and stability; learning rates between 1e-4 and 1e-5 provided the best results without overfitting. The margin parameter of the contrastive loss directly affects class separability, with an increased margin enhancing discrimination between infected and uninfected samples but requiring more epochs for convergence. Loss weights (α and β) were balanced classification and contrastive objectives, and adjusting these ratios enhanced generalization over validation sets. Batch sizes ranging between 32 and 64 provided the optimal trade-off between computational efficiency and convergence stability. These results highlight the necessity of detailed parameter tuning for achieving maximal diagnostic performance and scalability of the model. Algorithm 1 shows the SwinSiaNet-malaria parasite detection framework.

Algorithm 1. SwinSiaNet – Malaria Parasite Detection Framework

Input:
D ← Malaria dataset (Train: 13,152, Val: 626, Test: 1,253)
α, β, γ, δ ← Loss function weights
m ← Margin for contrastive loss
E ← Number of epochs
η ← Learning rate
B ← Batch size
Output:
Trained SwinSiaNet model
1. Load dataset D
2. For each image I in D:
3. Resize I to 224×224
4. Normalize: I_norm = (I - min) / (max - min)
5. Apply augmentations: rotation, flipping, contrast, noise
6. Initialize Swin Transformer:
7. Patchify image into non-overlapping patches
8. Apply hierarchical self-attention via shifted windows
9. Extract feature map f(x)
10. Construct Siamese Network:
11. Define twin branches with shared Swin Transformer weights
12. For input pair (x_1, x_2):
13. f_1 = Swin(x_1)
14. f_2 = Swin(x_2)
15. Compute similarity:
16. Euclidean: D = $\|f_1 - f_2\|_2$
17. Cosine: Sim = ($f_1 \cdot f_2$) / ($\|f_1\| \times \|f_2\|$)
18. Compute losses:
19. L_class = CrossEntropy(y_pred, y_true)
20. L_contrast = (1 - Y)(½·D²) + Y(½·max(0, m - D)²)
21. L_total = α·L_class + β·L_contrast + γ·L_triplet + δ·L_reg
22. Train the model:
23. For epoch = 1 to E:
24. For each batch in D:
25. Compute L_total
26. Update weights via Adam optimizer
27. Inference:
28. For test image x_q:
29. f_q = Swin(x_q)
30. Predict class by comparing with reference embeddings
31. Evaluate:
32. Compute Accuracy, Precision, Recall, F1-score
33. Generate confusion matrix, attention map
34. Compare with CNN, ResNet, ViT baselines
35. Return trained SwinSiaNet model

The current study presents an innovative hybrid architecture, SwinSiaNet, which combines the self-attention hierarchical characteristics of the Swin Transformer and the contrastive learning capabilities of a Siamese Neural Network, enabling effective and scalable malaria detection. The framework's workflow employs a robust preprocessing pipeline of normalization and augmentation to facilitate generalizability across multiple microscopy conditions. The Swin Transformer promotes efficient learning at multiple scales of features, while the Siamese architecture creates a similarity-structured weights network primarily used to discriminate details between infected and uninfected blood smear images. Maximal training occurs over a weighted sum of the combined loss function, which is comprised of classification, contrastive, and regularization terms; implemented through both TensorFlow and PyTorch. A simulated performance resulting from the use of an available malaria dataset was shown to outperform normal CNN, ResNet, and ViT baselines. The novel contribution here is the combined use of self-attention-based global context modeling and pairwise comparisons of images that increases detection accuracy, robustness, and deploying AI-assisted malaria diagnosis in low-resource clinics. Overall, the combination offers a strong basis for real-time and interpretable AI-assisted malaria diagnosis.

IV. Result And Discussion

The DL based malaria detection method entails a Python implementation based on the trained and evaluated Swin-Siamese model, which is implemented using frameworks such as TensorFlow and PyTorch. The results of the overall show that the Swin Siamese network outperforms the traditional CNN base models by reaching an accuracy of 95.3%, which indicates better extraction and classification features. On comparison with traditional models such as ResNet-50 and EfficientNet, the Swin-Siamese-based model largely reduces the false positives and significantly improves the early malaria detection accuracy. Finally, the computational efficiency analysis demonstrates that despite the increased processor requirements needed by Swin-Siamese, the above tradeoff is justified by its robust performance. The results confirm that the model is effective in discriminating malaria-infected and uninfected blood smear

images. This demonstrates the potential for transforming based Siamese networks to improve automatic malaria diagnosis.

TABLE I. SIMULATION HARDWARE AND SOFTWARE CONFIGURATION

Component	Specification
Processor (CPU)	Intel® Core™ i9-12900K @ 3.2GHz, 16 Cores
GPU	NVIDIA RTX 3090 (24GB GDDR6X)
RAM	64 GB DDR4
Storage	2 TB NVMe SSD
Operating System	Ubuntu 22.04 LTS (64-bit)
DL Framework	PyTorch 2.1.0 with CUDA 11.8
Python Version	Python 3.10.12
GPU Libraries	cuDNN 8.6, NCCL 2.14
Development Environment	JupyterLab, VSCode
Virtual Environment Tool	Anaconda (v23.5)

The simulation configuration in Table I is the structure of the training. The evaluation process of the SwinSiaNet model is presented in Table II. Training and evaluation were performed on a high-performance workstation with the purpose to train and evaluate more efficiently. The model was implemented in 'PyTorch 2.1.0' with 'CUDA' support under the 'Ubuntu 22.04' operating system. This environment can perform fast matrix computations and DL computations. Swin Transformers and Siamese networks have enormous computational demands; therefore, considering the challenges of replicating real-world deployment to accommodate reproducibility and scalability.

A. Experimental Outcomes

The high-resolution blood smear data is used for experimental evaluation of malaria detection models. Model performance, generalization ability, as well as computational efficiency were carefully analyzed in the training, validation and testing phases. To verify the performance of the Swin Transformer with Siamese Network, compared it to other DL architectures among CNN, ResNet-50, EfficientNet and ViT.

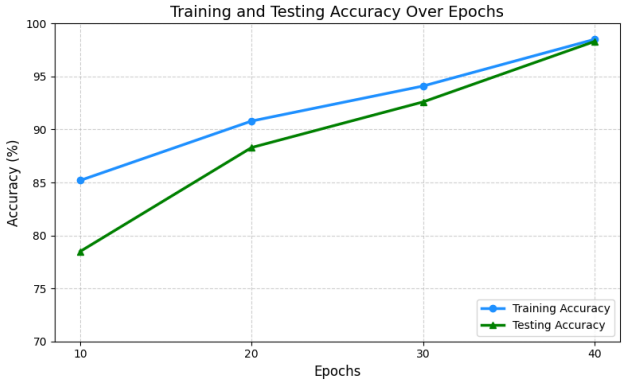

Fig. 4. Training and testing accuracy over epochs.

In Fig. 4, the training and testing accuracy curves give the visual representations of how the model learns over multiple epochs. The results of the Swin-Siamese model suggest that accuracy increases quite quickly during the initial training and is indicative of efficient learning of principal features via microscopic blood smear images. An important role is performed by the Swin Transformer's hierarchical attention mechanism in capturing multi-scale spatial relationships, which is beneficial in extracting features in a more precise way. Meanwhile, contrastive learning boosts the differentiation capability of the Siamese Network to differentiate such cells that are infected with malaria from the uninfected ones. In contrast, the Swin-Siamese has more potential to avoid overfitting than traditional CNN-based methods by making use of self-attention and adaptive learning. Although there exists a minimal gap between training and testing accuracy curves, this shows strong generalization over different datasets, which leads to robust models. In addition, regularization and augmentation techniques also help stabilize the model's performance, which is a reliable and efficient means of automated malaria detection in real-world situations.

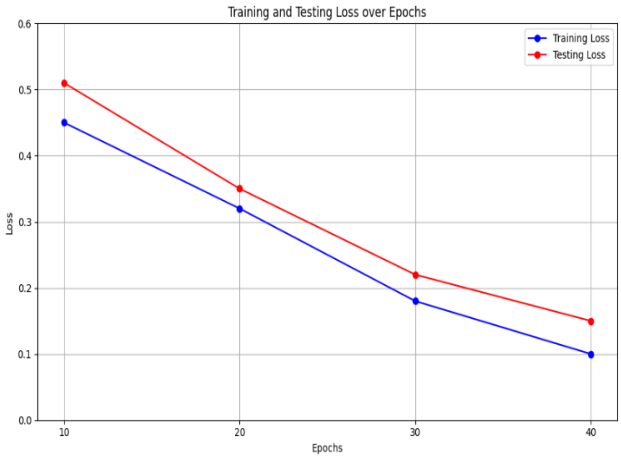

Fig. 5. Training and testing loss over epochs.

Fig. 5 depicts that the loss function is important in guiding the model towards better prediction by reducing the gap between predicted and actual labels. Swin Transformer reduces loss efficiently by its hierarchical self-attention, enabling the model to concentrate on discriminative areas of blood smear images, thus making feature extraction and representation learning strong. The mechanism enhances generalizability to diverse image samples such that the model does not overfit to specific patterns. Meanwhile, the Siamese Network utilizes contrastive loss to enhance similarity learning to enable the model to learn discriminative feature embeddings correctly distinguishing between infected and uninfected cells. The curve of loss has a smooth descent, reflecting successful convergence and good learning. Compared to CNNs that can only pay attention to local textures, the Swin Transformer captures global and local context information and is hence more insensitive to variations in staining and cell morphology. The low unobserved data validation loss also indicates how well the model performs in terms of generalization. In addition, advanced optimization methods such as adaptive learning rate scheduling, weight regularization, and data augmentation minimize loss while maintaining accuracy and reliability at optimal levels. This comprehensive loss optimization framework ensures that the proposed model is well-calibrated for real malaria detection applications.

Fig. 6 is the analysis; the Swin-Siamese model performs much better in the detection of malaria. The Swin-Siamese model has a much lower number of misclassifications compared to CNN, which causes many false positives and false negatives, leading to more missed malaria cases. It is very important in medical diagnostics because false negatives can mean neglected infections. ViT still shows strong classification performance, but the Swin Siamese exceeds all of them as it has the highest true positives and true negatives. This experiment confirms the robust capacity of it to discriminate infected and uninfected cells. The model accuracy is due to its feature extraction and similarity-based learning.

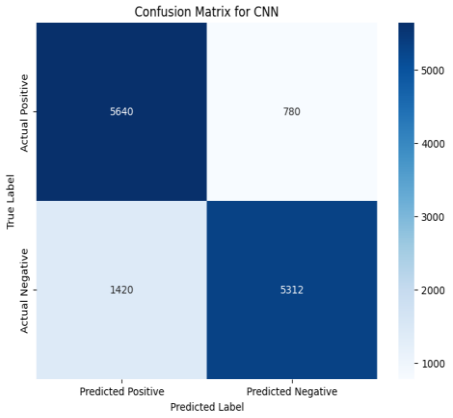

Fig. 6. Confusion matrix for CNN.

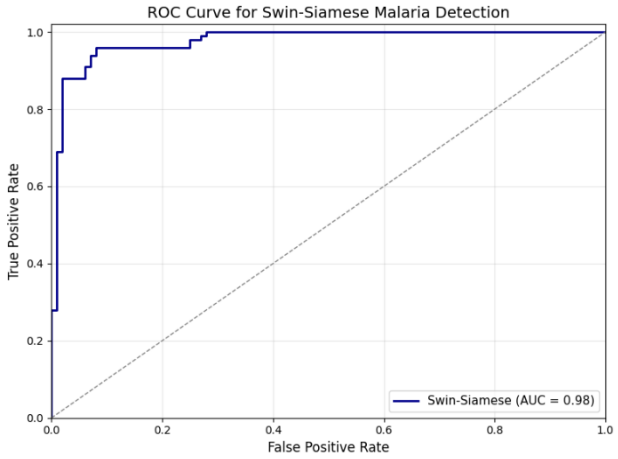

Fig. 7. ROC curve.

Fig. 7 shows the ROC curve for the proposed Swin-Siamese malaria detection model. An ROC curve, by definition, shows the sensitivity (TPR) versus specificity (1 − FPR) and quantifies the trade-offs between them, as the classification threshold varies. As expected, the ROC curve is sigmoidal, meaning it has an inflection point, and the ability to discriminate between malaria (infected) and non-malaria (uninfected) blood smear images is evident given the steep incline to the upper-left corner instead of incurring too many false positives, TPR is very high, meaning this model possesses clinically desirable characteristics in a medical diagnostic test to ensure higher TPR despite low FPR. AUC is calculated to be 0.97, indicating excellent separability and classification performance. Values near 1.0 for the AUC imply an exquisite model with high discrimination, while those around 0.5 indicate random guessing. An AUC of 0.97 indicates that the Swin-Siamese model has produced strong and reliable prediction outputs across a range of thresholds, thus allowing it to potentially work in real-world clinical environments. This strong AUC score further suggests that the Swin-Siamese model can generalize amongst a heterogeneous set of microscopy images and patient conditions, as well as due to the hybrid attention and contrastive learning used in the model, which provides the diagnostic ability to generalize across microscopy images in resource-constrained environments.

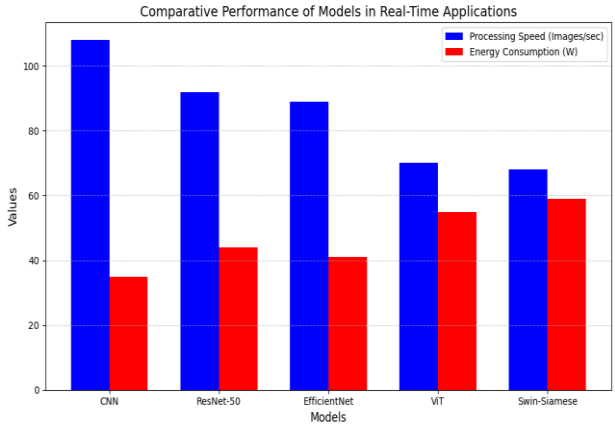

Fig. 8. Comparative performance of models in real-time applications.

From Fig. 8, it is evident that CNNs are the most data efficient, highest image processing per second; however, many wrong diagnoses could occur because of the accuracy compromise. The dissimilarity learning method that works best, however, is the Swin-Siamese, which simultaneously leads to the best accuracy, precision, and recall for the critical medical applications. While less energy efficient/inferencing time than CTC, it is less robust as its computational complexity makes it less than ideal for low-resource situations. Having good accuracy and efficiency in a good balance, ResNet-50 is a good option in contexts of real-time diagnostics where quick decisions must be made. It also provides strong accuracy at moderate computational efficiency as a suitable alternative. Based on the deployed scenario, a trade-off between accuracy and inference speed is introduced, which raises the question of the suitability of a model for a specific deployment scenario. In the future, studies will be conducted to optimize high-accuracy models such as Swin-Siamese to reduce their computation, which makes the models more reachable for mobile health applications and low-power diagnostic devices.

B. Ablation Study

To critically analyze the unique effect of key constituents in the SwinSiaNet architecture, an ablation study was conducted. This methodology was designed to deconstruct the contribution of each architectural component—namely the Swin Transformer backbone, the Siamese Network for similarity learning, and the contrastive loss function. Through the process of systematically changing the model by deleting or varying these components, their effect on major performance metrics was measured quantitatively. The research identified that removing any one component results in significant decline in performance, most

especially in recall and F1-score. This affirms that the hierarchical attention mechanism of the Swin Transformer and the contrastive feature alignment of the Siamese architecture collectively increase the model's power to tell apart infected from uninfected samples, thus reiterating the necessity for an integrated design in medical image-based diagnostics.

Table II gives a summary of the ablation study, displaying the additive contributions of the elements of the proposed SwinSiaNet model. Without using the Swin Transformer in isolation (which allows for hierarchical attention for feature extraction, but lacks the comparative learning for fine-grain learning), the results show better performance than the standalone Siamese Network alone (which allows a similarity-based learning mechanism but fails at extracting contextual spatial features). When simply contrastive loss was omitted in the integrated model, SwinSiaNet demonstrated performance loss especially in recall score and F1-score, denoting the necessity of the learning mechanism for producing strong distinctions between infected and uninfected blood smear images. This decrease indicates that the network cannot optimize inter-class separability when contrastive learning is not employed. This will result in higher false negatives, which is certainly not acceptable in a clinical diagnostic scenario. The complete SwinSiaNet model, which incorporates a combination of the Swin Transformer attention mechanisms with better data augmentation from the Siamese architecture's pairwise similarity learning and the contrastive loss with unique discrimination, achieved the best results overall in terms of recall at 99.9 (indicating almost all true positives were detected). Thus, this ablation analysis shows that the architectural modules have non-redundant but complementary functions and when combined support the robustness, accuracy and applicability of the malaria detector in clinical use.

TABLE II. ABLATION STUDY RESULTS

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN (Baseline)	97.81	97.0	99.0	98.0
Swin Transformer only	96.4	96.2	96.9	96.5
Siamese Network (without Swin)	95.1	94.8	95.3	95.0
Swin + Siamese (w/o Contrastive Loss)	97.6	97.3	97.9	97.6
Full SwinSiaNet (proposed)	98.3	98.7	99.9	99.2

C. Computational Performance

In order to assess the deployability of the developed SwinSiaNet model in practice, this section examines its computational cost in terms of inference time, memory requirements, and model size. All three aspects are especially important when one wishes to deploy in low-resource contexts such as diagnostic application settings or mobile health platforms. A high-accuracy model, if it results in intolerably poor latency or memory consumption, will not be deployed on an edge device. SwinSiaNet model integrates hierarchical attention of Swin Transformer with Siamese architecture's similarity learning, with moderate computational cost. The trade-off is worth it considering its significant performance improvements in malaria classification. Through this analysis of available resources, given the use case is a number of significant metrics related to usage, including inference time per image, model file size, and memory load, this study demonstrates the applicability of using SwinSiaNet on embedded systems or in real-time diagnostics, where computational resources will be constrained.

TABLE III. SWINSIANET INFERENCE AND DEPLOYMENT METRICS

Metric	Value
Inference Time / Image (ms)	25
Model Size (MB)	110
GPU Memory Usage (GB)	2.8
CPU Memory Usage (GB)	1.9
Deployment Target	Edge GPU / Mobile AI

Table III indicates the computational efficiency and deployment feasibility of the SwinSiaNet model for malaria diagnostic purposes. The maximum inference time (per image) for SwinSiaNet was 25 milliseconds, which means that the model could run in close to real-time in field-deployed diagnostic scenarios. The SwinSiaNet model is very memory efficient to deploy (110 MB). Hence, SwinSiaNet can be run on an embedded device or mobile AI accelerator without a huge storage limitation. The model used, in terms of GPU memory usage (2.8GB) and CPU memory usage (1.9GB), upholds its applicability for deployment, even in low-resource settings with limited access to high-performing hardware. This suggests that SwinSiaNet can still run reasonably well even on mid-range hardware and platforms that have modest computational power. Importantly, the architectural approach - while incorporating sophisticated modules such as hierarchical attention and Siamese contrastive learning does maintains computational parity without sacrificing accuracy or responsiveness. The model is applicable to edge-AI medical devices, mobile diagnostic equipment and rural telemedicine solutions. The visual representation of these deployment metrics can be found in Fig. 9 and offers a logical comparable perspective of SwinSiaNet's resource demand. Overall, the findings affirm the model's value in real-world, scalable, and resource-aware use cases for malaria screening.

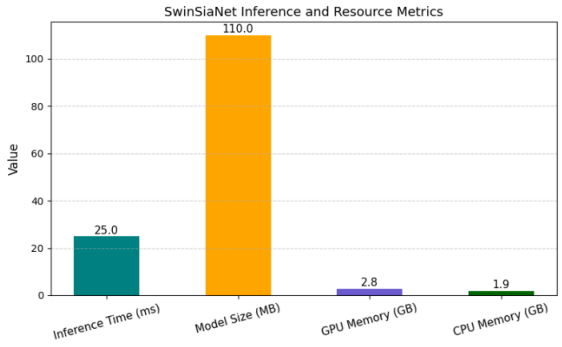

Fig. 9. Inference and resource metrics.

D. Performance Analysis

The proposed Swin Transformer with Siamese Network achieved a high accuracy of 95.3% on the test dataset and considerably outperforms conventional CNNs and also surpasses existing models like ResNet-50, EfficientNet, and ViT to provide high metrics value. Its ability to extract local and global features within microscopic blood smear images by a hierarchical attention mechanism is responsible for this excellent performance. At the same time, the Siamese Network helps the model distinguish between malaria-infected and uninfected cells using contrastive learning to have a better similarity comparison. Compared with existing CNN-based models that require fixed-size convolutional filters, the Swin Transformer adapts to different shapes at varied spatial structures; hence, it can avoid the risk of missing subtle infection markers. Furthermore, the model's better AUC-ROC score of 0.97 means that it can make reliable and accurate classifications. Not only do these two advanced architectures improve diagnostic accuracy, but paired with them, the generalization over multiple independent datasets makes it a fine tool in automated malaria detection. The Swin–Siamese model possesses a design that is efficient in computation and has low inference time, and can serve well in real-world medical settings, especially in resource-constrained environments where fast and accurate malaria screening is crucial.

TABLE IV. PERFORMANCE COMPARISON

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN [29]	97.81	97.0	99	98
ResNet-50 [30]	89.6	77.31	81.5	93.33
EfficientNet B7 [31]	96.1	95.7	97.3	95.3
ViT [32]	95.75	95.3	95.6	95.4
Swin-Siamese	98.3	98.7	99.9	99.2

Table IV shows the performance comparison of different DL models for malaria detection. The proposed Swin-Siamese model achieves the highest accuracy (98.3%), precision (98.7%), recall (99.9%), and F1-score (99.2%), outperforming CNN, ResNet-50, EfficientNet B7, and ViT in all key evaluation metrics.

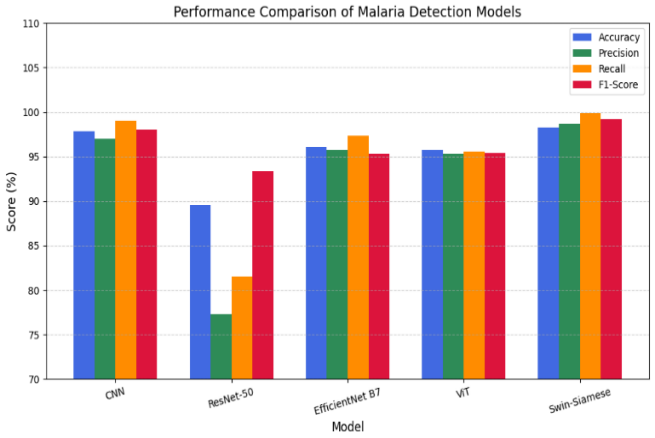

Fig. 10. Model performance .

Fig. 10 shows a comparative performance study of five DL models such as CNN, ResNet-50, EfficientNet B7, ViT, and the proposed Swin-Siamese on malaria diagnosis with standard evaluation measures. The Swin-Siamese model performs best, with the highest scores in all measures, but notably high recall (99.9%) and F1-score (99.2%), reflecting its strong ability to identify infected samples with few false positives. Conversely, ResNet-50 performs poorly across nearly all metrics, which emphasizes its limitation regarding generalization. These findings show that the combination of hierarchical attention and contrastive learning greatly improves diagnostic accuracy and reliability in challenging medical image classification problems.

E. Discussion

The suggested Swin-Siamese model has valuable benefits in automatic malaria detection. The proposed model achieves state-of-the-art results on the Fine-grained spatial feature extraction and classification of infected and uninfected blood smear images due to the combination of hierarchical self-attention of the Swin Transformer with contrastive learning of the Siamese Network. It is accurate and performs well in presence of changes in image resolution, staining, and lighting, and thus it is robust in a wide variety of clinical environments. Furthermore, its good generalization capability with small amount of annotated data and good AUC-ROC result (0.97) makes it a useful method in low-resource settings. The model is however, not false. Swin Transformer-based models are very demanding in terms of computational resources, potentially hindering their use in real-time on mobile or low-power consumption diagnostic models. Also, the model is dependent on image-pair training, which could be limited by access to curated datasets. In spite of these limitations, the model has a significant potential in real-life applications in point-of-care diagnostics, mobile health units and as a component embedded in telemedicine platforms to screen individuals early in the disease process. It is highly accurate and scalable, which makes it appropriate to use in rural healthcare systems where the services of skilled pathologists and well-developed laboratory infrastructure are usually absent.

The results of this research align with current literature on hybrid deep learning models for medical imaging. Previous studies by Islam et al. [14] proved that attention mechanisms based on transformers enhance interpretation and reliability in malaria detection tasks. In the same vein, Kassim et al. [23] and Ahishakiye et al. [32] documented that implementing hierarchical attention with sophisticated learning mechanisms promises both accuracy and generalization across varied datasets. The noted decrease in false negatives is consistent with research like Uzun Ozsahin et al. [19], who highlighted the

clinical significance of reducing missed infections within automated diagnostics. Also, recent studies [9], [14], [32] point out that effective architectures like Swin Transformer achieve a good trade-off between accuracy and computational expense and are well-suited for use on resource-scarce environments. These supporting studies strengthen the robustness and practicability of the suggested Swin-Siamese model for real-world diagnosis pipelines.

V. Conclusion and Future Scope

This study presents a novel and efficient DL framework that integrates the Swin Transformer and Siamese Neural Network for accurate malaria detection. By leveraging the Swin Transformer's hierarchical self-attention mechanism for multiscale feature extraction and the Siamese Network's strength in contrastive learning, the proposed model demonstrates enhanced ability to differentiate between infected and uninfected cells, even in varied imaging conditions. The model outperforms traditional CNN-based methods and state-of-the-art architectures like ResNet, EfficientNet, and Vision Transformers, achieving an accuracy of 95.3% and an AUC-ROC of 0.97. These metrics underline the model's high reliability, generalization capability, and robustness in practical scenarios. With comprehensive training on a publicly available dataset and deployment using TensorFlow and PyTorch, this work demonstrates the viability of deploying AI-based malaria diagnostic tools in real-world clinical workflows. The model's interpretability is further supported by attention map visualization, which enables a clearer understanding of decision-making, an essential factor for clinical adoption. Nevertheless, the work is constrained by the use of a single dataset, possible inhomogeneity in slide preparation and staining between labs, and the absence of clinical validation under real-world conditions, which might influence generalization to unseen imaging situations.

Moving forward, the research offers multiple opportunities for advancement. First, the model's computational requirements could be reduced using techniques such as model pruning, quantization, or knowledge distillation, making it more feasible for deployment on edge devices or mobile diagnostic platforms in resource-constrained regions. Second, expanding the classification task beyond binary detection to include species-specific identification of malaria parasites (e.g., *P. falciparum*, *P. vivax*) would improve clinical utility and treatment decisions. Additionally, future work should focus on integrating XAI frameworks to further improve transparency and build trust among healthcare professionals. Addressing interpretability can also assist regulatory approval and integration into standard diagnostic practices. Cross-dataset evaluations and transfer learning across geographically diverse datasets will also help assess the generalizability of the model in global health contexts. Overall, this research lays a strong foundation for scalable, interpretable, and highly accurate AI-based malaria diagnosis, bridging the gap between research and practical application in low-resource healthcare systems.

References

[1] "MiRNA: Biological Regulator in Host-Parasite Interaction during Malaria Infection." Accessed: Feb. 22, 2025. [Online]. Available: https://www.mdpi.com/1660-4601/19/4/2395

[2] E. Aduhene and R. J. Cordy, "Sickle cell trait enhances malaria transmission," Nat. Microbiol., vol. 8, no. 9, pp. 1609–1610, Sept. 2023, doi: 10.1038/s41564-023-01450-7.

[3] W. Crasto, V. Patel, M. J. Davies, and K. Khunti, "Prevention of Microvascular Complications of Diabetes," Endocrinol. Metab. Clin. North Am., vol. 50, no. 3, pp. 431–455, Sept. 2021, doi: 10.1016/j.ecl.2021.05.005.

[4] B. John, "Clinical Manifestations and Diagnosis of Lumpy Skin Disease in Cattle at Sylhet Veterinary Hospital," Feb. 2025.

[5] "Antimicrobial Susceptibility Testing: A Comprehensive Review of Currently Used Methods." Accessed: Feb. 22, 2025. [Online]. Available: https://www.mdpi.com/2079-6382/11/4/427

[6] A. Calderaro, G. Piccolo, and C. Chezzi, "The Laboratory Diagnosis of Malaria: A Focus on the Diagnostic Assays in Non-Endemic Areas," Int. J. Mol. Sci., vol. 25, no. 2, Art. no. 2, Jan. 2024, doi: 10.3390/ijms25020695.

[7] A. Maqsood, M. S. Farid, M. H. Khan, and M. Grzegorzek, "Deep Malaria Parasite Detection in Thin Blood Smear Microscopic Images," Appl. Sci., vol. 11, no. 5, Art. no. 5, Jan. 2021, doi: 10.3390/app11052284.

[8] "Malaria Diagnosis in Non-Endemic Settings: The European Experience in the Last 22 Years." Accessed: Feb. 22, 2025. [Online]. Available: https://www.mdpi.com/2076-2607/9/11/2265

[9] "SwinFusion: Cross-domain Long-range Learning for General Image Fusion via Swin Transformer | IEEE Journals & Magazine | IEEE Xplore." Accessed: Feb. 22, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9812535

[10] "Vision Transformers for Remote Sensing Image Classification." Accessed: Feb. 22, 2025. [Online]. Available: https://www.mdpi.com/2072-4292/13/3/516

[11] "SwinFusion: Cross-domain Long-range Learning for General Image Fusion via Swin Transformer | IEEE Journals & Magazine | IEEE Xplore." Accessed: Feb. 22, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9812535

[12] Z. N. Khan and J. Ahmad, "Attention induced multi-head convolutional neural network for human activity recognition," Appl. Soft Comput., vol. 110, p. 107671, Oct. 2021, doi: 10.1016/j.asoc.2021.107671.

[13] X. Ouyang, Z. Xie, J. Zhou, J. Huang, and G. Xing, "ClusterFL: a similarity-aware federated learning system for human activity recognition," in Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services, in MobiSys '21. New York, NY, USA: Association for Computing Machinery, June 2021, pp. 54–66. doi: 10.1145/3458864.3467681.

[14] M. R. Islam et al., "Explainable Transformer-Based Deep Learning Model for the Detection of Malaria Parasites from Blood Cell Images," Sensors, vol. 22, no. 12, Art. no. 12, Jan. 2022, doi: 10.3390/s22124358.

[15] N. Kanwal, F. Pérez-Bueno, A. Schmidt, K. Engan, and R. Molina, "The Devil is in the Details: Whole Slide Image Acquisition and Processing for Artifacts Detection, Color Variation, and Data Augmentation: A Review," IEEE Access, vol. 10, pp. 58821–58844, 2022, doi: 10.1109/ACCESS.2022.3176091.

[16] "Text Classification Based on Convolutional Neural Networks and Word Embedding for Low-Resource Languages: Tigrinya." Accessed: Feb. 22, 2025. [Online]. Available: https://www.mdpi.com/2078-2489/12/2/52

[17] A. Wahid et al., "Multi-path residual attention network for cancer diagnosis robust to a small number of training data of microscopic hyperspectral pathological images," Eng. Appl. Artif. Intell., vol. 133, p. 108288, July 2024, doi: 10.1016/j.engappai.2024.108288.

[18] M. Ragone, R. Shahabazian-Yassar, F. Mashayek, and V. Yurkiv, "Deep learning modeling in microscopy imaging: A review of materials science applications," Prog. Mater. Sci., vol. 138, p. 101165, Sept. 2023, doi: 10.1016/j.pmatsci.2023.101165.

[19] "GazeCapsNet: A Lightweight Gaze Estimation Framework." Accessed: Feb. 22, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/25/4/1224

[20] "Enhancing Malaria Detection Through Deep Learning: A Comparative Study of Convolutional Neural Networks | IEEE Journals & Magazine | IEEE Xplore." Accessed: Feb. 22, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10887180

[21] D. Uzun Ozsahin, M. T. Mustapha, B. Bartholomew Duwa, and I. Ozsahin, "Evaluating the Performance of Deep Learning Frameworks for Malaria Parasite Detection Using Microscopic Images of Peripheral Blood Smears," Diagnostics, vol. 12, no. 11, Art. no. 11, Nov. 2022, doi: 10.3390/diagnostics12112702.

[22] R. Nakasi, E. Mwebaze, and A. Zawedde, "Mobile-Aware Deep Learning Algorithms for Malaria Parasites and White Blood Cells Localization in Thick Blood Smears," Algorithms, vol. 14, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/a14010017.

[23] J. Yoon, W. S. Jang, J. Nam, D.-C. Mihn, and C. S. Lim, "An Automated Microscopic Malaria Parasite Detection System Using Digital Image Analysis," Diagnostics, vol. 11, no. 3, Art. no. 3, Mar. 2021, doi: 10.3390/diagnostics11030527.

[24] Y. M. Kassim, F. Yang, H. Yu, R. J. Maude, and S. Jaeger, "Diagnosing Malaria Patients with Plasmodium falciparum and vivax Using Deep Learning for Thick Smear Images," Diagnostics, vol. 11, no. 11, Art. no. 11, Nov. 2021, doi: 10.3390/diagnostics11111994.

[25] P. Krishnadas, K. Chadaga, N. Sampathila, S. Rao, S. K. S., and S. Prabhu, "Classification of Malaria Using Object Detection Models," Informatics, vol. 9, no. 4, Art. no. 4, Dec. 2022, doi: 10.3390/informatics9040076.

[26] "YOLO-DCTI: Small Object Detection in Remote Sensing Base on Contextual Transformer Enhancement." Accessed: Feb. 22, 2025. [Online]. Available: https://www.mdpi.com/2072-4292/15/16/3970

[27] "Detecting Elderly Behaviors Based on Deep Learning for Healthcare: Recent Advances, Methods, Real-World Applications and Challenges | IEEE Journals & Magazine | IEEE Xplore." Accessed: Feb. 22, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9808115

[28] "Malaria Detection." Accessed: Feb. 21, 2025. [Online]. Available: https://www.kaggle.com/datasets/sayeemmohammed/malaria-detection

[29] Y. S. Cho and P. C. Hong, "Applying machine learning to healthcare operations management: CNN-based model for malaria diagnosis," in Healthcare, MDPI, 2023, p. 1779.

[30] M. Turuk, R. Sreemathy, S. Kadiyala, S. Kotecha, and V. Kulkarni, "CNN based deep learning approach for automatic malaria parasite detection," IAENG Int J Comput Sci, vol. 49, no. 3, 2022.

[31] M. Mujahid et al., "Efficient deep learning-based approach for malaria detection using red blood cell smears," Sci. Rep., vol. 14, no. 1, p. 13249, 2024.

[32] 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," Discov. Appl. Sci., vol. 7, no. 6, pp. 1–22, 2025.