

CHAPTER I

INTRODUCTION

1.1. Background

Malaria is a potentially fatal infectious disease caused by *Plasmodium* parasites and remains a significant public health challenge in numerous endemic countries, including Indonesia. The nation's tropical environmental conditions, characterized by high temperatures, abundant rainfall, and favorable habitats for *Anopheles* mosquitoes, support the continuous transmission of the disease. Based on data published by the Indonesian Ministry of Health, Indonesia recorded 543,965 malaria cases in 2024, with nearly 90% of the reported infections originating from Papua Province. Globally, According to the World Health Organization (WHO), malaria affects more than 200 million individuals each year, underscoring its persistent impact on global health. The disease remains particularly challenging in regions where diagnostic resources are limited and the availability of skilled microscopists is insufficient. Although microscopic examination of stained blood smears is regarded as the reference standard for malaria diagnosis, this approach presents several limitations. Its diagnostic accuracy largely depends on the proficiency of laboratory personnel, the examination process is labor-intensive and time-consuming, and the likelihood of diagnostic inaccuracies increases when analyzing blood smears with low parasite density or suboptimal image quality. [1].

Recent developments in artificial intelligence have significantly promoted the widespread adoption of deep learning techniques to improve the speed and accuracy of malaria detection using microscopic blood smear images. Among these techniques, Convolutional Neural Networks (CNNs) have become the most widely adopted approach in both national and international studies. A wide range of studies has shown that CNN-based models attain high performance in identifying *Plasmodium*-infected blood cells. Study [1] reported a classification accuracy exceeding 90% for malaria detection using CNNs, while study [2] achieved an accuracy of 98.30%. Similarly, study [3] which utilized a large-scale

dataset comprising 27,558 microscopic images, reported an accuracy of 96% using a CNN-based model. Furthermore, a systematic review conducted in [4] revealed that the majority of CNN-based malaria detection studies consistently achieved classification accuracies above 90% across both national and international datasets, highlighting the effectiveness of CNNs for malaria classification tasks. In addition, model interpretability has increasingly been incorporated into recent studies. Study [5] demonstrated that integrating saliency mapping with CNNs improved classification accuracy to 95.06% by enabling the model to focus on the most informative regions of microscopic images. The findings reported in both [6] dan [5] further confirm the capability of CNNs to effectively extract discriminative features from microscopic blood smear images.

Despite these promising results, the majority of existing studies continue to rely on conventional CNN architectures, which exhibit inherent limitations in capturing global spatial relationships across image regions. By design, CNNs operate using local receptive fields through convolutional kernels, making them less effective at modeling long-range spatial dependencies and the overall structural context of an image. This limitation becomes particularly significant in microscopic malaria images, which often exhibit substantial variations in staining intensity, slide artifacts, inconsistent parasite sizes, and visual noise. Consequently, there is an increasing need for more adaptive models capable of handling the high degree of heterogeneity commonly encountered in real-world microscopic image datasets. In Indonesia, research on deep learning-based malaria detection has predominantly focused on CNN architectures, with limited exploration of more recent Transformer-based models. This is despite the fact that, from a theoretical perspective, Transformer architectures offer superior capabilities for capturing global spatial dependencies and contextual information within images.

Recent advances in computer vision have demonstrated a paradigm shift from conventional CNN-based architectures toward Transformer-based models, particularly the Vision Transformer (ViT) and Swin Transformer. These architectures employ a self-attention mechanism to model global relationships across image regions, enabling the extraction of richer and more contextual feature

representations than those produced by traditional CNNs. Study [7] reported that the Swin Transformer, through its shifted-window mechanism, exhibits greater stability when processing high-resolution and heterogeneous images, including medical images. Furthermore, study [8] demonstrated that the Vision Transformer and its variants outperform conventional CNN architectures across a wide range of medical image classification tasks by generating more informative feature representations. In addition, study [9] showed that the Swin Transformer is capable of capturing blood cell patterns more effectively, indicating its strong potential for microscopic blood image analysis, including malaria diagnosis.

International datasets, such as those provided by the National Institutes of Health (NIH) and Kaggle, are extensively utilized in malaria research due to their high-quality annotations, large sample sizes, and representative image diversity. The morphological characteristics of erythrocytes and the visual appearance of Plasmodium parasites are generally consistent across populations, making these international datasets highly relevant for research conducted in the Indonesian context. Studies [10] and [11] demonstrated that models trained on NIH and Kaggle datasets achieved high classification performance, even when applied to research conducted in national settings. Moreover, with appropriate image preprocessing and data augmentation techniques, these global datasets can be effectively generalized to microscopic images obtained from Indonesian laboratories, despite variations in sample preparation procedures, staining intensity, and slide quality.

Furthermore, binary malaria classification (infected versus uninfected) plays a fundamental role in clinical practice, as it serves as the initial step in determining the presence of parasitemia and assessing the severity of infection. Accurate identification of infected erythrocytes is essential because manual microscopic examination is often challenged by variations in blood smear quality, low levels of parasitemia, and the subjective interpretation of laboratory personnel. Previous studies conducted in Indonesia have demonstrated that automated image-based classification systems provide more consistent diagnostic assessments than conventional manual examinations, thereby reducing the likelihood of diagnostic errors [4], [12]. Moreover, several studies have emphasized that the development

of automated malaria detection systems is particularly valuable in healthcare facilities with limited access to trained microscopists. Consequently, deep learning-based malaria classification systems have the potential to serve as reliable clinical decision support tools by providing faster, more objective, and standardized diagnostic outcomes [13].

Based on the existing body of literature, several important research gaps can be identified. First, to the best of the author's knowledge, no study conducted in Indonesia has investigated the application of Vision Transformer (ViT)-based models, particularly the Swin Transformer architecture, for automated malaria classification. Second, previous studies have not explored a hybrid framework that combines the capability of Convolutional Neural Networks (CNNs) to capture fine-grained local features with the Swin Transformer's strength in modeling long-range contextual dependencies. Third, the majority of existing research has focused primarily on classification accuracy as the main performance indicator, while comprehensivesuch as precision, recall, the F1-score, and the area under the receiver operating characteristic (ROC) curve (AUC), have received comparatively less attention despite their importance in assessing diagnostic reliability in medical image analysis. Finally, the integration of Explainable Artificial Intelligence (XAI) approaches, particularly Gradient-weighted Class Activation Mapping (Grad-CAM), remains limited, reducing the interpretability of classification results and the ability to identify biologically meaningful image regions that contribute to the model's decision-making process.

The novelty of this study lies in the development of a Hybrid CNN–Swin Transformer architecture for binary malaria classification (infected versus uninfected), the comprehensive evaluation of model performance using multiple clinically relevant metrics, and the integration of Grad-CAM to improve the transparency and interpretability of the proposed model. By combining the complementary strengths of CNNs and Transformer-based architectures, this research aims to develop a more robust and reliable automated malaria detection system. Ultimately, the proposed approach is expected to contribute significantly to the advancement of artificial intelligence-based malaria diagnosis by providing

a solution that is more accurate, adaptive, interpretable, and suitable for practical clinical implementation.

1.2. Research Problems

Based on the foregoing discussion, the research questions investigated in this study are formulated as follows:

1. How can a Hybrid CNN–Swin Transformer model for binary malaria classification be designed and implemented to effectively combine the local feature extraction capability of CNNs with the global attention mechanism of the Swin Transformer?
2. In what way can Explainable Artificial Intelligence (XAI) through Gradient-weighted Class Activation Mapping (Grad-CAM) be employed to improve the interpretability of Plasmodium-infected blood cell classification by revealing the image regions that drive the model's prediction?
3. To what extent does the proposed Hybrid CNN–Swin Transformer model outperform the individual CNN and Swin Transformer architectures when evaluated using clinically relevant performance measures, including accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC)?
4. How can the proposed classification model be incorporated into a web-based prototype to facilitate rapid, accurate, and user-friendly malaria detection?

1.3. Research Objectives

The present study is intended to develop and assess a Hybrid CNN–Swin Transformer model for the binary classification of malaria in microscopic blood cell images, distinguishing between infected and uninfected cells. The proposed architecture combines the ability of CNNs to capture discriminative local features with the Swin Transformer's capacity to learn global spatial dependencies from microscopic images. Model performance is comprehensively evaluated using clinically relevant metrics, and its transparency is further improved by incorporating Explainable Artificial Intelligence (XAI) techniques, specifically

Gradient-weighted Class Activation Mapping (Grad-CAM), to generate visual interpretations of the model's predictions.

1.4. Research Significance

This research is expected to contribute in several ways

1. For the Researcher:
 - A. To provide practical experience in designing, training, and comparing modern deep learning architectures, particularly the Hybrid CNN–Swin Transformer model, as well as applying Explainable Artificial Intelligence (XAI) techniques.
 - B. To enhance expertise in medical image analysis, including image preprocessing, data augmentation, hyperparameter tuning, and model performance evaluation using multiple clinically relevant metrics.
 - C. To develop practical skills in building a web-based medical detection system that can be implemented in real-world healthcare applications.
2. For Healthcare Professionals and Clinical Laboratories:
 - A. To provide an AI-assisted diagnostic tool capable of analyzing microscopic malaria images more rapidly, accurately, and consistently.
 - B. To reduce reliance on manual microscopic examination, which is susceptible to subjectivity, particularly in healthcare facilities with limited numbers of trained microscopists.
 - C. To improve the transparency of the proposed model by generating interpretable visual explanations through Gradient-weighted Class Activation Mapping (Grad-CAM), thereby strengthening clinicians' confidence in its diagnostic predictions.
3. For Researchers and Readers:
 - A. To contribute to the existing body of knowledge on the development of Hybrid CNN–Transformer models for medical image classification, with a particular focus on malaria diagnosis.
 - B. To provide a comprehensive comparison of the performance of CNN, Swin Transformer, and Hybrid CNN–Swin Transformer architectures in binary image classification.

- C. To establish a foundation for future research on automated diagnostic methods for other tropical infectious diseases.

1.5. Scope and Limitations

To ensure that the study remains focused and well-defined, the following limitations are established:

1. Classification: This study is limited to binary classification of microscopic malaria images, consisting of two classes: infected and uninfected.
2. Models: The study's scope is limited to three architectures, namely Convolutional Neural Networks (CNN), Swin Transformer, and the proposed Hybrid CNN–Swin Transformer model.
3. Interpretability: Model interpretability is assessed exclusively using Gradient-weighted Class Activation Mapping (Grad-CAM).
4. Dataset: This study employs the publicly available NIH Malaria Dataset accessed through Kaggle, which consists of 27,558 microscopic blood cell images, including 13,779 infected samples and 13,779 uninfected samples. The original dataset labels are retained throughout the study, and no additional annotation or relabeling procedures are applied.
5. Data Splitting: The dataset is partitioned into training, validation, and testing sets with a ratio of 80%, 10%, and 10%, respectively, using a stratified sampling approach to ensure that the class distribution remains balanced across all subsets.
6. System Development: The proposed system is implemented as a web-based prototype intended for research purposes and does not encompass clinical validation or integration with hospital laboratory information systems.