Computers in Biology and Medicine 186 (2025) 109704

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A deep architecture based on attention mechanisms for effective end-to-end detection of early and mature malaria parasites in a realistic scenario

Luca Zedda *, Andrea Loddo *, Cecilia Di Ruberto

Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124, Cagliari, Italy

ARTICLE INFO

Keywords:
Computer vision
Deep learning
Image processing
Malaria parasites detection
Early malaria diagnosis

ABSTRACT

Background: Malaria is a critical and potentially fatal disease caused by the Plasmodium parasite and is responsible for more than 600,000 deaths globally. Early and accurate detection of malaria parasites is crucial for effective treatment, yet conventional microscopy faces limitations in variability and efficiency.
Methods: We propose a novel computer-aided detection framework based on deep learning and attention mechanisms, extending the YOLO-SPAM and YOLO-PAM models. Our approach facilitates the detection and classification of malaria parasites across all infection stages and supports multi-species identification.
Results: The framework was evaluated on three publicly available datasets, demonstrating high accuracy in detecting four distinct malaria species and their life stages. Comparative analysis against state-of-the-art methodologies indicates significant improvements in both detection rates and diagnostic utility.
Conclusion: This study presents a robust solution for automated malaria detection, offering valuable support for pathologists and enhancing diagnostic practices in real-world scenarios.

1. Introduction

Malaria is a critical and potentially fatal disease caused by the Plasmodium parasite. It is primarily transmitted through the bites of infected female Anopheles mosquitoes. According to WHO's 2023 World Malaria Report, an annual assessment of global trends in malaria control and elimination, the estimated number of global malaria cases in 2022 exceeded pre-COVID-19 pandemic levels in 2019. The report highlights several threats to the global response to malaria, including climate change [1].

In 2022, there were an estimated 249 million reported cases and 608,000 deaths occurring globally. Of the 249 million cases noted in 2022, 233 million (around 94%) were in the WHO African Region. More than 50% of all deaths occurred in just four countries—Nigeria (31%), the Democratic Republic of the Congo (12%), Niger (6%), and Tanzania (4%). Around 70% of the global malaria burden is concentrated in 11 countries belonging to the African region, referred to as "High Burden to High Impact" (HBHI); in 2022, there were an estimated 167 million cases (67% of the global total) and 426,000 deaths (73% of the global total) in the original HBHI countries [1].

The Plasmodium parasites that cause malaria in humans include five species: Plasmodium falciparum (Pf), Plasmodium vivax (Pv), Plasmodium ovale (Po), Plasmodium malariae (Pm), and Plasmodium knowlesi (Pk), with Pf and Pv posing the most significant threat [2,3]. The life cycle of the malaria parasite within the human host involves several distinct stages, including the ring, trophozoite, schizont, and gametocyte stages. Understanding these different stages is crucial for developing effective treatments and prevention strategies.

The WHO characterizes human malaria as both preventable and treatable if diagnosed promptly. Delayed diagnosis can lead to severe complications, including disseminated intravascular coagulation, tissue necrosis, and splenic hypertrophy [2,4]. Thus, accurate and timely diagnosis is essential for effective treatment.

Malaria diagnosis can be achieved through various techniques, including microscopic blood smear analysis, rapid diagnostic tests (RDTs), and real-time polymerase chain reaction (PCR). These methods are crucial, as malaria symptoms can often be misinterpreted as those of other diseases, such as viral hepatitis or dengue fever [5]. While PCR offers heightened accuracy, it is not always practical in programmatic settings due to its requirement for specialized laboratory facilities. Consequently, the WHO recommends that all suspected malaria cases be confirmed through microscopy or RDT prior to treatment, acknowledging the risks associated with false-negative results.

Microscopy remains the preferred diagnostic method among pathologists [6–8], particularly in endemic regions, due to its sensitivity, cost-effectiveness, and capacity to identify specific parasite species and densities [5,9,10]. The traditional methodology involves analyzing a

* Corresponding authors.
E-mail addresses: luca.zedda@unica.it (L. Zedda), andrea.loddo@unica.it (A. Loddo).

https://doi.org/10.1016/j.compbiomed.2025.109704
Received 19 August 2024; Received in revised form 8 January 2025; Accepted 14 January 2025
Available online 26 January 2025

peripheral blood smear (PBS) on a glass slide to detect malaria parasites and their developmental stages. However, this technique is not without limitations. Challenges include: *(i)* difficulty in detecting infections during early stages, necessitating skilled microscopists; *(ii)* a shortage of qualified microscopists, substandard quality control, and misdiagnosis due to low parasitemia or mixed infections in some endemic regions; *(iii)* limited access to microscopy in rural health facilities; *(iv)* challenges in accurately identifying different Plasmodium species, which can lead to misdiagnosis and implications for epidemiological understanding [9,10]; *(v)* technical expertise required for slide preparation; *(vi)* potential lysis of red blood cells, altering parasite morphology and complicating identification; *(vii)* variability in microscope quality and illumination; *(viii)* variability in staining procedures; *(ix)* influence of parasitemia levels on detection accuracy [5].

Furthermore, controlling infectious diseases remains imperative, particularly in underdeveloped countries lacking adequate medical infrastructure [4].

Accurate and timely diagnosis of malaria is vital for effective treatment and prevention of severe complications. While conventional microscopy is still regarded as the gold standard, advancements in deep learning (DL) techniques, particularly Convolutional Neural Networks (CNNs), have shown promise in enhancing malaria cell image analysis.

Recent studies have successfully applied CNNs for single-cell malaria diagnosis, underscoring the importance of accurately determining whether a cell is infected [11–13]. However, reliance on monocentric cell image datasets assumes an idealized scenario where salient and highly discriminative features are easily extracted from the images. This ideal condition is typically achieved through preliminary steps of detection or segmentation of the full-size images.

In practical applications, however, systems are fully automated, and the images may not always be perfectly centered or accurately cropped. These variations can lead to suboptimal detection and, consequently, reduced diagnostic precision. Nevertheless, previous research has demonstrated the effectiveness of detection systems in real-world scenarios for computer-aided diagnosis (CAD) systems, particularly those that are robust enough to accommodate these image quality challenges [14–18].

Nevertheless, additional challenges remain, including the differentiation of various Plasmodium species and addressing complexities associated with low parasitemia levels and asymptomatic infections. Consequently, achieving precise localization of parasites within cells is crucial for thorough investigations and accurate diagnostics [8,19,20].

In response to these challenges, we propose a novel CAD approach to automated early and already in-progress identification of malaria parasites and quantification of parasitemia. This dual-purpose system seeks to aid pathologists while addressing the limitations associated with conventional microscopy.

Our contributions to the field of malaria parasite detection are outlined as follows:

(i) We extend YOLO-SPAM [21] and YOLO-PAM [22] by proposing a novel deep learning-based detection framework that provides a more comprehensive approach to facilitate the detection of malaria parasites across all stages of infection, from early to mature. By incorporating specialized attention-based heads, our model significantly improves the recognition of small ring-stage parasites critical markers for early-stage infection surpassing previous state-of-the-art methods.

(ii) Our method can accurately perform multi-species detection on four distinct species of malaria, thereby accommodating both mixed infections and intra-species detection scenarios. By demonstrating strong performances across species, our models could be applied in the future due to the unavailability of such annotated data, to effectively detect multiple species in blood smear images

(iii) We introduce the ability to classify the various life stages of the four malaria species, enhancing the diagnostic utility of our approach in an end-to-end framework. By leveraging color and geometric-based augmentations we mitigate the stage class imbalance.

(iv) We conduct a comprehensive evaluation of our method across three publicly available datasets, providing a comparative analysis against state-of-the-art (SOTA) methodologies in the field. Our novel contribution improves drastically upun the performances of full CNN based methodologies [23] up to 23% in Average Precision

Using the You Only Look Once (YOLO) architecture, which has shown remarkable results in our prior research [21,22,24], we have implemented enhancements to improve its accuracy for detecting small-to-mature parasites. This enables not only early but also late infection diagnosis, as well as multi-species detection and life-stage classification of malaria parasites. The choice of the YOLO architecture, combined with strategically positioned attention mechanisms and a limited number of trainable parameters, enables our proposed models to be effectively deployed in low-resource settings. Notably, our largest model, YOLO Para AP, requires only 47,2 additional milliseconds per image compared to the YOLOv8 medium model.

The remainder of this article is organized as follows. First, a background about the task at hand and an overview of the literature is given in Section 2. Then, materials and methods are described in Section 3, while Section 4 describes the proposed architectures. The experimental evaluation and the obtained results follow in Section 5. A detailed discussion and a description of the limitations are given in Section 6. Finally, conclusions are drawn in Section 7.

2. Background

CAD systems in medicine extend beyond specific domains, finding significant applications in hematology. Numerous CAD solutions have been proposed for the automated detection of malaria parasites, which mitigate manual analysis errors and provide consistent interpretations of blood samples. This technological advancement ultimately leads to reduced diagnostic costs and improved efficiency in healthcare delivery [7,20].

The research of CAD methods is not limited to hematology but spans different fields such as COVID detection [53,54], detection of membrane proteins [55,56], or sensing in MRI [57]. The integration of CAD technology into hematology aims to enhance both the accuracy and speed of diagnoses, thereby improving patient outcomes and optimizing healthcare systems [58–63]. The current landscape of malaria parasite detection encompasses both traditional image processing methods and advanced DL techniques. Traditional approaches typically involve the detection or segmentation of parasites, feature extraction, and subsequent classification, either as independent tasks or as an interconnected pipeline. In contrast, end-to-end deep learning frameworks consolidate these processes, a shift driven by innovations such as AlexNet [64].

Conventional methodologies have employed mathematical morphology techniques for preprocessing and segmentation [14,15]. Additionally, handcrafted feature extraction has been utilized to train machine learning classifiers [17]. Over the past decade, DL approaches have gained prominence, with a substantial body of literature demonstrating their efficacy in improving detection accuracy and efficiency compared to traditional methods [8,11,13,20,65].

The existing literature on malaria parasite analysis from blood smear images can be categorized into four primary areas:

- parasite detection and classification from full-size images (Section 2.1);
- parasite classification from single-cell images (Section 2.2);
- domain generalization methods from high- to low-cost devices (Section 2.3);

Table 1
The table provides a comprehensive summary of the studies identified in the literature, including the authors, publication year, specific task faced, techniques employed, datasets used, number of images, and performance measures. These measures are abbreviated with mAP for mean Average Precision, A for Accuracy, SA for Segmentation Accuracy, F1 for F1-score, and DSC for Dice Similarity Coefficient.

Authors	Task	Methods	Dataset	Images	Performance (%)
Zhao et al. [25] (2020)	Pf detection	SSD300	BBBC041v1 [27]	1364	mAP: 90.40
Chibuta and Acar [26] (2020)		Modified YOLOv3	Dataset A [26]	2703	mAP: 88.00
			Quin et al. [28]	1182	mAP: 90.20
Abdurahman et al. [29] (2021)		Modified YOLOv4	Quin et al. [28]	1182	mAP: 96.32
Zedda et al. [23] (2022)		YOLOv5	MP-IDB [30]	104	mAP: 87.2
Zhao et al. [25] (2020)	Pf classification	VGG-16	NIH [12]	27,558	A: 96.53
Setyawan et al. [31] (2022)		Morphology and texture	MP-IDB [30]	250	A: 82.67
Zedda et al. [23] (2022)		DarkNet-53	MP-IDB [30]	1297	F1: 95.58
Penas et al. [32] (2017)	Pf and Pv species detection	InceptionV3	Private	363	A: 87.90
Maity et al. [8] (2020)	Pf and Pv segmentation	Capsule Network	MP-IDB [30]	210	SA: 99.10
			Private	38	SA: 98.70
Rahman et al. [33] (2021)	Differentiate healthy and infected RBCs	Custom CNN	BBBC041v1 [27]	1364	A: 99.35
			NIH [12]	27,558	A: 99.35
			MP-IDB [30]	229	F1: 84.82
Loh [34] (2021)		Mask R-CNN	BBBC041v1 [27]	15,144	A: 94.57
Li et al. [35] (2021)		CNN DTGCN	BBBC041v1 [27]	15,144	A: 98.50
Acherar et al. (2020) [36]		VGG-19	Private	1250	A: 99.70
		EfficientNet-B7	NIH [12]	27,558	A: 98.80
Silka et al. [37] (2022)		Custom CNN	BBBC041v1 [27]	1364	A: 99.68
Kumar et al. [38] (2024)		Hybrid Capsule Network	NIH [12]	27,558	A: 99.07
Krishnadas et al. [39] (2022)	Detect the four malaria parasite species	Scaled YOLOv4	MP-IDB [30]	172	mAP: 83.0
Mukherjee et al. [40] (2021)		Custom CNN	MP-IDB [30]	210	DSC: 95.0
Nautre et al. [41] (2022)		U-Net	MP-IDB [30]	21	A: 99.4
Nanoti et al. [42] (2016)	Classify the four malaria parasite species	SVM	Private	300	A: 90.17
Var and Tek [43] (2018)		VGG-19	Private	654	A: 87.50
Abbas [44] (2020)		Random Forest	Malaria-Detection-2019 [44]	263[a]	A: 82.00
Chaudhry [45] (2024)		Custom CNN	MP-IDB [30]	275	A: 96.00
			IML [46]	86	A: 92.00
			Malaria-Detection-2019 [44]	263[a]	A: 82.00
Hung and Carpenter [47] (2017)	Classify the parasite's life stages	Faster R-CNN	BBBC041v1 [27]	1364	A: 72.00
Manku et al. [48] (2020)		Faster R-CNN	BBBC041v1 [27]	560	A: 82.00
Loddo et al. [49] (2022)		CNN	MP-IDB [30]	140	A: 99.40
Arshad et al. [46] (2021)		ResNet50v2	IML with RBC [46]	669	A: 95.63
Sengan et al. [50] (2022)		ViT	IML [46]	669	A: 90.03
Li et al. [51] (2021)	Classify WBCs vs. healthy RBCs vs. Pv life stages	Custom CNN	BBBC041v1 [27]	1364	A: 98.30
Yang et al. [52] (2020)	Pv detection	Cascaded YOLO	Private	2567	mAP: 79.22
Sultani et al. [20] (2022)		Faster R-CNN	M5 [20]	1257	mAP: 66.80

[a] Indicates that only ring, trophozoite, and schizont classes are available.

- methods for low-cost sensor image devices (Section 2.4).

Table 1 provides a comprehensive summary of the most relevant works in the several tasks of malaria parasites analysis.

2.1. Parasite classification from full-size images

Detecting malaria parasites from blood smear images presents considerable challenges, particularly in settings with limited clinical resources, such as underdeveloped countries [17]. The analysis of full-size images is essential for near real-time diagnosis, but distinguishing parasites from similarly structured cell components, such as white blood cells and platelets, complicates identification. Microscopic evaluation of peripheral blood smears can take over 15 min per slide, and recognizing the various life cycle stages of Plasmodium adds another layer of complexity [66].

A meticulous analysis is vital for accurate diagnoses of conditions like malaria and leukemia, necessitating robust segmentation [67] and detection techniques to delineate regions of interest (ROIs) before classification [68]. Some studies focus directly on classifying full-size images using CNNs or traditional machine learning techniques trained on either handcrafted features or those extracted from pre-trained CNNs. For instance, Vijayalakshmi et al. [6] employed a Support Vector Machine (SVM) trained with features from a VGG-19 network to differentiate infected from non-infected malaria images.

Recent advancements have introduced DL methods that streamline the multi-stage pipeline. Arshad et al. [65] proposed a framework that combines U-Net segmentation with watershed algorithms, followed by binary classification to identify healthy versus infected cells, and further classification of the life cycle stages of the infected cells, utilizing ResNet50v2. Similarly, Maity et al. [8] implemented a semantic segmentation approach followed by a Capsule Network for classifying Plasmodium falciparum rings. Conversely, Sultani et al. [20] compared various off-the-shelf object detectors, including Faster R-CNN, RetinaNet, and YOLO, specifically for the life stages of P. vivax.

Additionally, Lin et al. [35] and Manescu et al. [69] developed custom object detection pipelines aimed at malaria diagnosis.

Our research has identified four additional studies employing YOLO for parasite detection, including YOLOv3 and YOLOv4 for thick blood smears [26,29,70] or thin ones [71].

The current state of the art is increasingly incorporating attention-based architectures across various fields, including malaria detection. Fu et al. [72] demonstrated that attention mechanisms, combined with self-supervised learning, can substantially enhance performance in malaria detection. Similarly, works such as [21,22] have shown that integrating attention mechanisms within the YOLO architecture improves detection accuracy. These benefits are not limited to malaria identification [73]; Zhu et al. [74] illustrated how these methods are effective for detecting small objects in drone imagery. Despite the performance gains offered by attention mechanisms over traditional CNN architectures, these techniques often come with significant memory demands, especially in the case of self-attention [75], which scales quadratically with input size, making naive implementations unsuitable for resource-limited environments. Lightweight attention mechanisms, such as Convolutional Block Attention Module and Normalized Attention Module [76,77], are computationally efficient and require fewer trainable parameters, making them more viable for low-resource settings. Typically, low-resource attention mechanisms are applied to feature maps with larger spatial dimensions, while more complex mechanisms can be used in later stages of computation near the prediction heads [21,22].

2.2. Parasite classification from single-cell images

Given that malaria parasites primarily affect red blood cells (RBCs), methods targeting the classification of individual cells have emerged, aiming to differentiate between parasitized and healthy erythrocytes [11–13,33,38,78]. These studies often utilize the NIH dataset [12] for benchmarking. Recent investigations have begun to explore the application of vision transformer within the same dataset, exemplified by Sengar et al. [79], reflecting a growing interest in the potential of transformers in deep learning-based malaria research.

Specific solutions include ad-hoc designed CNN architectures [11], transfer learning on CNNs pre-trained on ImageNet, such as ResNet-50 [12] and VGG-19 [33], and ensemble approaches combining VGG-19 with SqueezeNet [13]. Notably, Diker et al. [78] introduced a residual CNN architecture optimized through Bayesian methods to extract critical features, subsequently inputting these features to an SVM classifier.

2.3. Domain generalization methods from high- to low-cost devices

In the realm of computer-aided medical image analysis, machine learning techniques frequently encounter the domain shift issue arising from discrepancies between source and target data distributions. Domain adaptation has garnered attention as a viable solution to this challenge [80,81].

Sultani et al. [20] addressed the difficulties of acquiring images in resource-limited areas by compiling a dataset using both low-cost and high-cost microscopy. They evaluated various domain adaptation techniques, aiming to effectively apply high-cost microscope images as the source domain and low-cost images as the target.

Further exploration of domain adaptation tasks remains necessary, such as the capability to classify different Plasmodium species based solely on knowledge of one species (e.g., recognizing P. malariae, P. vivax, and P. ovale when only P. falciparum data is available).

Contrastive learning techniques have shown strong performances in mitigating the variability across high-to-low-cost devices, Dave et al. [82] show this leveraging domain adaptive contrastive loss as part of the training procedure.

2.4. Methods for low-cost sensor image devices

Low-cost mobile devices, including smartphones and tablets equipped with microscope cameras, have been increasingly leveraged for image acquisition and analysis. Applications specifically designed for smartphones, often utilizing pre-trained or customized CNNs and standard preprocessing techniques such as contrast enhancement [83], have been developed to facilitate automated malaria diagnosis [17,84], achieving impressive classification rates in as little as ten seconds [84].

The proliferation of affordable mobile devices has proven particularly beneficial in resource-limited countries, where high malaria mortality rates coincide with a lack of specialized diagnostic personnel and equipment [84]. This technological advancement offers a cost-effective solution for accurate malaria diagnosis [66].

2.5. Limitations of the existing literature

Despite the progress in analyzing full-size images, several limitations persist in the existing literature. Direct classification using CNNs or traditional machine learning approaches may oversimplify the diagnostic task, risking the loss of fine-grained details essential for accurate disease identification. Consequently, some studies resort to off-the-shelf object detectors on custom datasets. However, past research on malaria parasite detection has predominantly concentrated on thick blood smears, raising concerns about generalizability due to dataset and feature variations that can significantly impact model performance. Moreover, there is a noticeable gap in the analysis of multiple malaria species and life stages, as most studies tend to focus on a specific species.

Our investigation emphasizes the analysis of three publicly available thin blood smear image datasets. This strategic choice enables a detailed examination of fine-grained features within the images, facilitating the identification and detection of different species and life stages from both quantitative and qualitative perspectives.

Furthermore, as illustrated in Fig. 1, the three datasets exhibit distinct intrinsic characteristics, including variations in coloration, illumination conditions, composition, and parasite types. This diversity underscores the proposed method's ability to address the detection of small-to-large parasites and the identification and classification of their species and life stages within the broader context of malaria infections.

3. Materials and methods

The methodology proposed in this work is a deep learning-based framework to detect and classify malaria parasites in full-sized microscopic blood smear images. Specifically, our approach builds upon state-of-the-art object detection architectures, leveraging the YOLO framework with enhancements in terms of attention mechanisms for improved feature representation. Three publicly available datasets containing images of infected RBCs were used. This section describes the datasets employed in Section 3.1 and gives an overview of the state-of-the-art object detectors in Section 3.3. Furthermore, it delves into the concept of attention and its applications in the field (Section 3.4).

3.1. Datasets

This subsection analyzes and describes the datasets used in this work. Sections 3.1.1 to 3.1.3 report specific information for each dataset, including the instrumentation used to acquire the images, while Section 3.1.4 provides an overview of the differences existing in the different parasite life stages. Representative images of the different datasets are depicted in Fig. 1, while a brief comparative analysis to emphasize the distinctive characteristics is given in Table 2. Every dataset involved in this study comprises Giemsa-stained microscopic images.

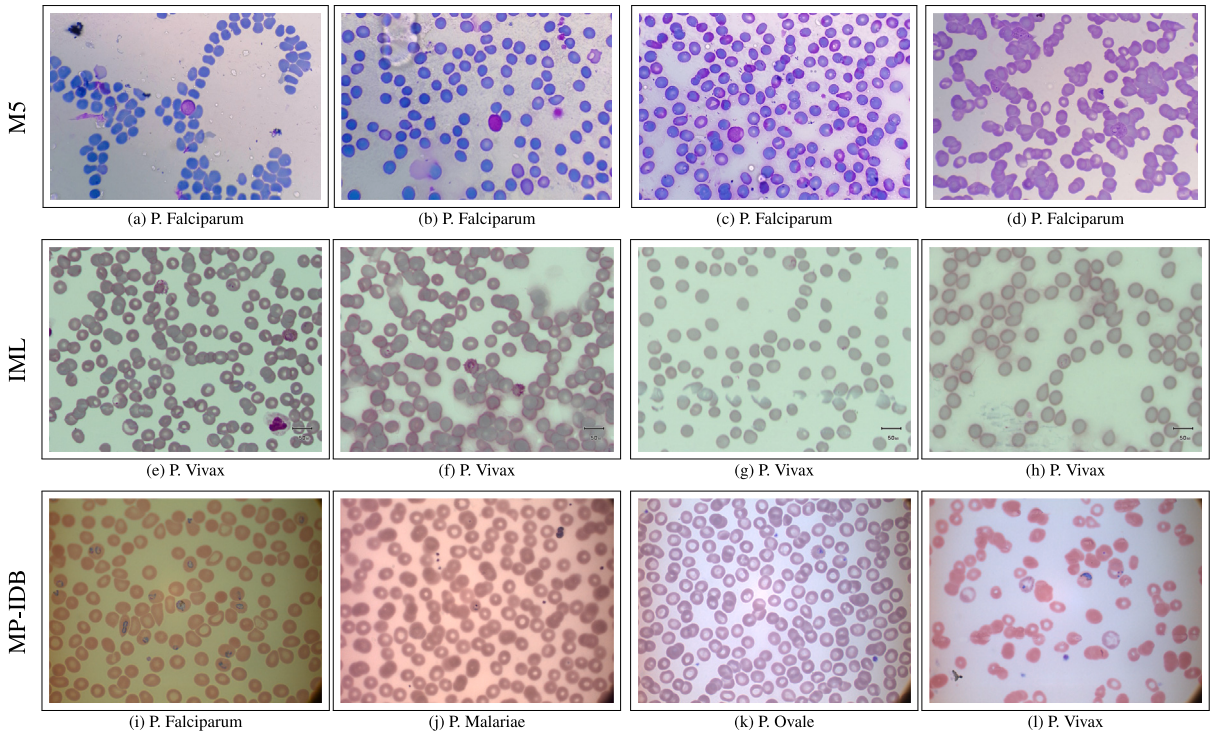


Fig. 1. Examples of full-size images from the M5, IML, and MP-IDB datasets. From top to bottom: Figs. 1(a) to 1(d) are samples from M5 (P. Falciparum only), Figs. 1(e) to 1(h) are from IML (P. Vivax only), and Figs. 1(i) to 1(l) are from MP-IDB (P. Falciparum, P. Malariae, P. Ovale, and P. Vivax, respectively).

Table 2
Comparison of the malaria datasets exploited in this study.

Dataset	Multi-stage	Samples	Species	Magnifications
M5 [85]	Yes	20,331	1	100×, 400×, 1000×
IML [46]	Yes	529	1	100×
MP-IDB [30]	Yes	172	4	–

3.1.1. M5

The M5 dataset [85], also known as the Multi Microscope Multi Magnification Malaria Dataset, contains 1257 images of thin blood smears. The same regions were monitored and captured on two different microscopes at three different magnifications (100×, 400×, 1000×) using a high-cost microscope and a low-cost microscope. The resolutions of the images are varied and not reported for simplicity.

3.1.2. IML

The IML dataset [46] comprises 345 images of blood samples from individuals infected with P. Vivax malaria in Pakistan's Punjab province. Each image contains an average of 111 blood cells and corresponding ground truth labels for the life stages and red blood cells. The images have a resolution of 1280×960 pixels and a 24-bit color depth and were taken using a microscope-attached camera magnified at 100×.

3.1.3. MP-IDB

The Malaria Parasite Image Database for Image Processing and Analysis (MP-IDB) [30] includes 210 images with a resolution of 2592× 1944 pixels of four types of malaria species - P. Falciparum, P. Malariae, P. Ovale, and P. Vivax - with each image corresponding to one or more of the four life stages of the species. The dataset features high-resolution images with a 24-bit color depth, allowing for detailed analysis of the variations within and across species.

3.1.4. Comprehensive datasets analysis with details on life stages

Each of the used datasets has the corresponding stage species for at least one represented species, as shown in Fig. 2.

Specifically, the ring stage is the first stage of the Plasmodium parasite's life cycle. After the parasite enters the human host, it invades liver cells and multiplies asexually, producing many ring-shaped structures. These rings are small and contain a single nucleus. Over time, the rings mature into other forms, such as the trophozoite and schizont stages.

The trophozoites are larger and more visible than the ring forms. They have a distinct nucleus and a more prominent cytoplasm. During this stage, the parasite continues multiplying asexually within the liver cells.

The schizonts are large, multinucleated structures that contain many daughter cells. When they are mature, they rupture the liver cells and release the daughter cells, called merozoites, into the bloodstream. The gametocytes are mature, non-dividing cells that do not replicate within the human host. Instead, they are taken up by a mosquito when it feeds on an infected human. Once inside the mosquito, the gametocytes can differentiate into male and female gametes, fertilizing each other to produce a zygote.

The different stages present various and, most importantly, hard-to-recognize characteristics. However, the main challenge related to the stage classification task is the high imbalance of the stages. This issue requires particular attention, often introducing proper image augmentation techniques for oversampling; another popular solution is to use imbalanced specialized classification algorithms and training processes. A complete representation of this issue is presented in Table 3. It provides the available count of each class across the different datasets. Please note that, for MP-IDB, only P. Falciparum is reported as the other species are not provided with their stage ground truth, i.e.,the label needed for the classification task.

3.2. Classification algorithms

For the sake of this work, we selected several off-the-shelf architectures to demonstrate their robust capabilities in the downstream

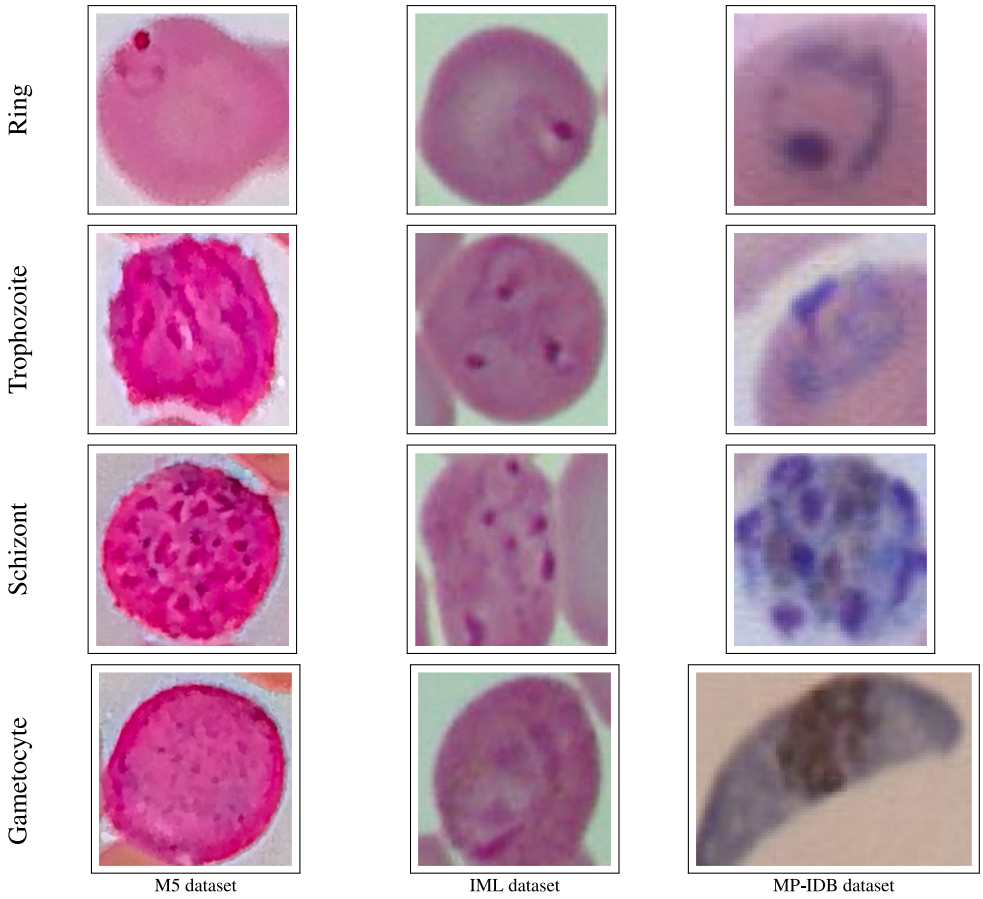


Fig. 2. Samples of various developmental stages of parasites as observed in distinct datasets. From top to bottom, the columns present instances from the M5, IML, and MP-IDB datasets, respectively. The last one specifically pertains to the P. falciparum species. The rows are organized to represent the developmental stages in the following sequence: ring, trophozoite, schizont, and gametocyte.

Table 3
Count of parasites present in the three datasets, based on ground truths provided by the authors.

Dataset	Species	Ring	Schizont	Trophozoite	Gametocyte
M5 [85]	P. Falciparum	2445	19	1088	90
IML [46]	P. Vivax	164	27	77	261
MP-IDB [30]	P. Falciparum	1230	18	42	7

task of malaria stage classification, especially given the limited existing literature on the application of such novel methods in this specific field. In addition, several common CNN architectures have been already considered in our previous work [24,49]. In particular, we selected two recent CNNs and three Vision Transformer (ViT)-based methods. We now provide a brief description of the architectures employed.

InternImage is a large-scale CNN that utilizes deformable convolutions to enhance its performance in tasks such as classification, detection, and segmentation. It has been shown to achieve accuracy on par with or better than ViTs while maintaining fewer parameters, making it a competitive alternative in large-scale vision tasks [86]. **ConvNeXt** is a modernized version of traditional CNNs, designed to bridge the gap between convolutional architectures and transformer models. It incorporates design principles from ViT while optimizing for performance in various vision tasks, demonstrating strong results in benchmarks [87].

DINO is a self-supervised learning framework that enhances the training of ViTs by leveraging knowledge distillation techniques without requiring labeled data. This approach allows for effective learning from unlabeled datasets, fostering robust feature extraction capabilities [88].

ViT revolutionized image processing by applying transformer architectures, traditionally used in natural language processing, to visual data. It excels in learning long-range dependencies within images but often requires larger datasets and more computational resources compared to CNNs [89].

Swin Transformers, introduced by Microsoft Research in 2021, apply transformer architecture to computer vision (CV) tasks [90]. These models process image patch embeddings using multi-headed self-attention modules, allowing for linear computational complexity with image size and enabling cross-window connections.

Swin Transformers employ hierarchical feature maps, similar to CNNs, downsampling images by 4×, 8×, and 16×. This approach facilitates tasks such as object detection and instance segmentation, potentially replacing convolution in vision tasks despite requiring more parameters.

Together, these models illustrate the evolving landscape of computer vision, where hybrid approaches and novel architectures continue to push the boundaries of performance and efficiency.

3.3. Object detectors

Contemporary object detectors are predominantly founded on deep learning methodologies and can be categorized into two primary types: one-stage and two-stage detectors.

Two-stage architectures, exemplified by Faster R-CNN [91], initially extract Regions of Interest (ROIs) and subsequently conduct classification and bounding box regression through a coarse-to-fine approach.

In contrast, one-stage detectors, such as SSD [92], FPN [93], and the YOLO family [94–97], generate bounding boxes and class predictions directly from the predicted feature maps utilizing predefined anchors.

One-stage detectors are characterized by their speed and compactness, rendering them particularly suitable for time-sensitive applications and computationally constrained edge devices [98,99].

The recent success of Transformer architectures in image recognition has spurred the development of several end-to-end Detection Transformers (DETRs). Despite their impressive recognition accuracy, DETRs face challenges related to their intricate architectures and slow convergence rates [98].

To address these limitations, this paper proposes a modified version of the one-stage detector YOLOv8 that aims to enhance both efficiency and accuracy in malaria parasites detection.

YOLO. The YOLO family of detectors diverges from the traditional two-step approach that relies on region-selection methods, instead employing an end-to-end differentiable network that integrates bounding box estimation with object identification. YOLO partitions the input image into $S \times S$ constant-size grids, with a CNN predicting bounding boxes and class labels for each grid. Bounding boxes with confidence scores exceeding a specified threshold are selected to identify objects within the image. The CNN performs a single pass to generate predictions, and following non-maximum suppression, it outputs the identified objects along with their corresponding bounding boxes, ensuring that each object is detected uniquely.

YOLOv8 the eighth version of the popular YOLO architecture proposed by the Ultralytics team encompasses a family of architectures and models for object detection that is pre-trained on the Common Objects in Context (COCO) dataset [100].

This family consists of five distinct models that share a common architectural framework but vary in terms of breadth, depth, and the number of trainable parameters. The models are designated as *YOLOv8n* (nano), *YOLOv8s* (small), *YOLOv8m* (medium), *YOLOv8l* (large), and *YOLOv8x* (extra-large), each pre-trained on images with resolutions of either 640×640 or 1280×1280 pixels. Notably, the number of trainable parameters for each model is as follows: YOLOv8n contains 3.2 million parameters, YOLOv8s has 11.2 million, YOLOv8m includes 25.9 million, YOLOv8l comprises 43.7 million, and YOLOv8x features 68.2 million parameters.

The architecture of YOLOv8 consists of three essential components, similar to those of other single-stage object detectors: the backbone, neck, and prediction head.

The backbone serves as a pre-trained network specialized in feature extraction from the input image. This process involves reducing the spatial resolution of the image while simultaneously enhancing the resolution of the extracted features.

The neck component amalgamates the extracted features and generates three distinct scales of feature maps, commonly referred to as feature pyramids. This design significantly enhances the model's capacity to generalize effectively across objects of varying sizes and scales.

Subsequently, the prediction head utilizes anchor boxes on the feature maps, facilitating the detection of objects based on the previously generated feature representations.

Similar to YOLOv5, the YOLOv8 architecture employs the CSPDarknet53 architecture with a Spatial Pyramid Pooling (SPP) layer [101] as its backbone, utilizes the Path Aggregation Network (PANet) [102] as the neck, and incorporates the YOLO detection head [94].

Despite the notable advancements in detection speed, it is widely recognized that YOLO architectures encounter difficulties in detecting small objects when compared to two-stage detectors [94,99]. This challenge has been addressed in this study, particularly in scenarios where the smallest parasites, such as the initial ring stages, are present. In this case, they may not be sufficiently large to be effectively detected by a conventional detector.

3.4. Attention mechanisms

In this section, we briefly define the concept of attention applied in CV tasks (refer to Section 3.4.1), and present the attention modules employed in this work (refer to Section 3.4.2).

3.4.1. Attention in computer vision

Attention mechanisms, inspired by human cognitive processes, have become pivotal in computer vision. These mechanisms selectively focus on salient input features, enhancing efficiency and accuracy in perceptual processing.

Among the different types of attention mechanisms the one that revolutionized mostly the computer vision field is the self attention proposed by Vaswani et al. [75] in which the attention function links queries with key–value pairs, producing outputs through weighted sums. Formally, given input features $x_1, x_2, \ldots, x_n$ and desired output y, attention is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

where Q, K, and V are query, key, and value matrices, respectively, and d_k is the key vector dimension [75].

3.4.2. Types of attention modules

Two prevalent attention mechanisms are spatial and channel attention [103,104]. Spatial attention identifies crucial image positions, while channel attention focuses on inter-channel feature relationships.

The Convolutional Block Attention Module (CBAM) combines these mechanisms sequentially, enhancing feature refinement [105].

The Normalized Attention Module (NAM) addresses varying dot-product attention scores by normalizing them, enhancing training stability [106].

3.5. Performance measures

This section presents the performance measures used to evaluate the detection and classification experiments in Sections 3.5.1 and 3.5.2, respectively.

3.5.1. Detection measures

Object detection methodologies are typically assessed using the mean *average precision* (mAP) metric and its various derivatives [107]. The concept of precision is grounded in the Intersection over Union (IoU) metric, which quantifies detection accuracy. Specifically, IoU is defined as the ratio of the area of overlap between the predicted bounding box and the actual object to the total area encompassed by both.

When the IoU exceeds a predetermined threshold, the detection is deemed correct and classified as a *true positive* (TP). Conversely, if the IoU falls below this threshold, the detection is categorized as a *false positive* (FP). Furthermore, if the model fails to identify an object that is present in the ground truth, this is referred to as a *false negative* (FN).

In terms of detection evaluation, *Precision* (PRE) is defined in Eq. (2):

$$PRE = \frac{TP}{TP + FP} \tag{2}$$

where:

- *TP* denotes the number of instances belonging to the *positive* class that have been accurately predicted;
- *FP* refers to instances where a non-existent object is incorrectly identified or an existing object is detected in an incorrect location;
- *FN* indicates instances where a ground truth bounding box is not detected.

In this study, the experimental evaluations were conducted using five variants of the mAP metric:

- $\mathbf{AP}_{0.50:0.95}$ is assessed using ten different IoU thresholds, ranging from 50% to 95% in increments of 5%;
- $\mathbf{AP}_{50}$ is evaluated at a single IoU threshold of 50%;
- $\mathbf{AP}_{s}$ represents the AP calculated for small objects, defined as those with an area of less than 32^2 pixels;
- $\mathbf{AP}_{m}$ denotes the AP calculated for medium objects, characterized by an area between 32^2 and 96^2 pixels;
- $\mathbf{AP}_{L}$ signifies the AP calculated for large objects, defined as those with an area greater than 96^2 pixels.

In order to provide a clear and fair comparison, we computed the $AP_{0.50:0.95}$ and AP_{50} using the official YOLOv8 framework [108] while for AP_S, AP_M, and AP_L we use the standard coco AP evaluation tool [109].

3.5.2. Classification measures

The classification performance is evaluated using several measures, including *accuracy*, *recall*, *precision*, and the *F1-score*.

In the following subsections, we provide straightforward definitions of these metrics as they pertain to binary classification problems, followed by their generalizations for multiclass scenarios.

Standard definitions for binary classification problems. An example, denoted as e, is characterized by a pair $\langle i, t\rangle$, where i represents a list of feature values and t denotes the assigned category (i.e.,the target category). A dataset D is defined as a collection of such examples. When the dataset D contains two target categories, it constitutes a binary classification problem. In this context, the categories are referred to as *negative* and *positive*.

To assess the performance of a binary classifier on the dataset D, each instance is labeled as either *negative* or *positive* based on the classifier's output. Depending on the classification outcome and the actual target value, an instance will contribute to one of the following counts:

- *True Negatives (TN)*: The number of instances belonging to the *negative* class that have been accurately predicted;
- *FP*: The number of instances belonging to the *negative* class that have been incorrectly predicted as positive;
- *FN*: The number of instances belonging to the *positive* class that have been incorrectly predicted as negative;
- *TP*: The number of instances belonging to the *positive* class that have been accurately predicted.

Based on these quantities, the measures to evaluate the classification performance can be defined as follows:

- *Accuracy* (ACC) — The ratio of correctly classified instances to the total number of instances:
$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$
- *Precision* — The fraction of positive instances that are correctly classified among all instances classified as positive:
$$PRE = \frac{TP}{TP + FP} \tag{4}$$
- *Recall* (REC) — It measures the classifier's ability to correctly identify the positive class, calculated against FN:
$$REC = \frac{TP}{TP + FN} \tag{5}$$
- *F1-score* (F1) — Defined as the harmonic mean of *precision* and *recall*:
$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{6}$$

Standard definitions for multiclass classification problems. As previously mentioned, the measures outlined can also be generalized for multiclass classification scenarios. A straightforward approach to achieve this is to calculate the metrics for each category using a One-vs-Rest (OvR) strategy. Following this process, the average value of each binary measure is computed, yielding an informative metric for the multiclass model.

Three distinct averaging methods can be employed: micro, macro, and weighted. In this study, macro averaging has been adopted.

In summary, for a classification problem involving K classes, the metrics with macro averaging are calculated as follows:

- *Macro Average Precision* (where P_k denotes the precision for class k):
$$MacroPrecision = \frac{\sum_{k=1}^{K} P_k}{K} \tag{7}$$
- *Macro Average Recall* (where SEN_k denotes the sensitivity for class k):
$$MacroRecall = \frac{\sum_{k=1}^{K} SEN_k}{K} \tag{8}$$
- *Macro Average F1-score* (where P and R denote the macro average precision and recall, respectively):
$$MacroF1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{9}$$

4. Proposed architectures

As discussed in Section 1, malaria is a life-threatening infectious disease that poses a significant threat to human life globally. The importance of accurate diagnosis and timely treatment cannot be overstated, as these are critical factors in reducing the high mortality rates associated with the disease. However, the current standard method of manually examining blood smears, which relies heavily on the expertise of skilled hematologists, is labor-intensive and susceptible to errors, particularly in resource-limited regions. This highlights the urgent need for more efficient and reliable diagnostic techniques.

In response to this need, this section presents the proposed methodologies for detecting and classifying malaria parasites in microscopic blood smear images. We have developed several innovative DL architectures tailored for the detection of malaria parasites in full-sized images, addressing the limitations of current diagnostic approaches.

The primary impetus for the development of these novel DL-based architectural models stems from the critical importance of early malaria detection for effective treatment and management. Accurate detection of malaria parasites across all their life stages is essential for proper diagnosis. This work proposes a novel pipeline designed to first detect parasites in their various sizes and stages and then classify them accordingly. Accurate identification is essential to ensure prompt treatment, which directly impacts patient outcomes.

To tackle the challenge of detecting parasites of different sizes even during the initial stages of infection, we have introduced three novel DL-based architectures: *YOLO Para SP* (described in Section 4.1), *YOLO Para SMP* (described in Section 4.2), and *YOLO Para AP* (described in Section 4.3).

These models have been customized and integrated into enhanced versions of the YOLOv8 framework with the following purposes:

- **YOLO Para SP**: optimized for detecting small parasites (ring stage).
- **YOLO Para SMP**: designed for both small and medium-sized parasites.
- **YOLO Para AP**: capable of detecting parasites of all sizes and stages.

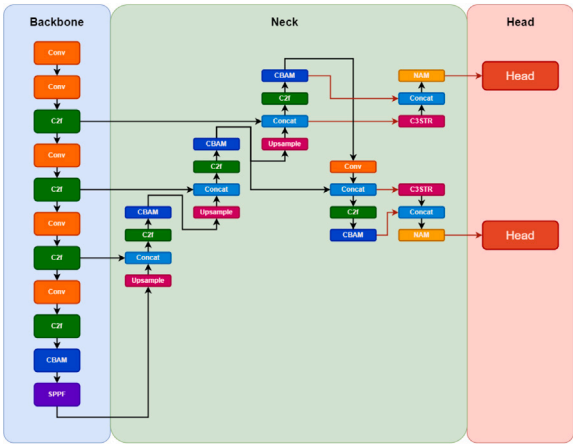


Fig. 3. YOLO Para SP architecture.

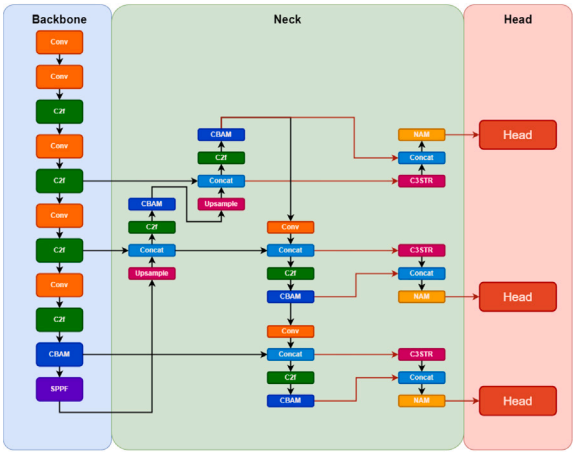


Fig. 4. YOLO Para SMP architecture.

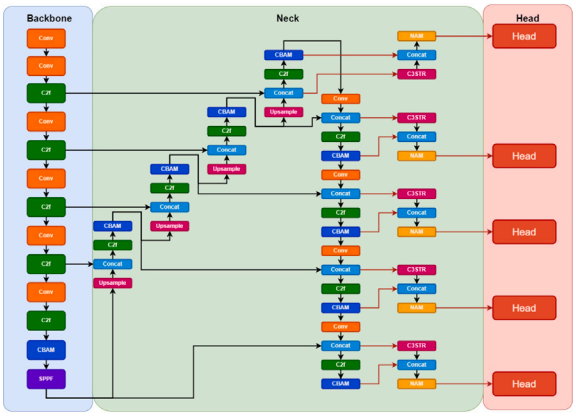


Fig. 5. YOLO Para AP architecture.

The **YOLO Para** series introduces several innovative enhancements specifically designed to improve the detection of parasites, with the ultimate goal of advancing malaria diagnostics. Each architecture incorporates multiple prediction heads, Swin Transformer-enhanced layers, and attention mechanisms to improve detection and classification accuracy, particularly for challenging cases such as low parasitemia or sources variability.

For clarity, the following abbreviations will be used throughout this text from now on: **SP** stands for *Small-sized Parasites*, **SMP** denotes *Small and Medium-sized Parasites*, and **AP** represents *All-sized Parasites*.

4.1. YOLO Para SP

The architectural design of *YOLO Para SP* is illustrated in Fig. 3. This model integrates *NAM* and *C3* modules, enhanced with Swin Transformer layers, into the YOLOv8 architecture, resulting in a highly advanced design optimized for detecting small parasites as shown by our previous work [21,22].

To boost the performance of *YOLO Para SP*, features extracted from the CBAM (Convolutional Block Attention Module) layers are fused with those derived from sequential C3STR (C3 with Swin Transformer) layers. This specialized feature fusion enhances the model's ability to detect small objects. Additionally, the NAM (Normalization and Attention Mechanism) module assigns higher weights to the most critical features while diminishing the importance of less relevant ones, thereby enriching the overall feature representation.

The lower layers of *YOLO Para SP* focus on extracting less refined features at a higher resolution, which is essential for detecting parasites that occupy only a few pixels in the image. Conversely, the higher-level layers excel in identifying medium-sized parasites but may lose some small parasite information due to the lower resolution of feature maps produced by convolution.

These enhancements in the *YOLO Para SP* architecture are aimed at improving the detection of parasites in full-size images, particularly focusing on the accuracy and efficiency of identifying small parasites.

4.2. YOLO Para SMP

Building upon the foundation laid by *YOLO Para SP*, the *YOLO Para SMP* architecture, shown in Fig. 4, introduces an additional prediction head, enhancing the model's capability to detect a broader range of parasite sizes. While retaining the core innovations of *YOLO Para SP*, such as the integration of NAM and C3STR modules, *YOLO Para SMP* is particularly optimized for detecting both small and medium-sized parasites. The additional prediction head is specifically designed to improve the model's sensitivity to medium-sized objects, ensuring a more comprehensive detection across different parasite scales. This makes *YOLO Para SMP* particularly effective in scenarios where parasites of varying sizes need to be identified simultaneously, enhancing the overall robustness and accuracy of malaria diagnostics.

4.3. YOLO Para AP

The *YOLO Para AP* model, illustrated in Fig. 5, extends *YOLO Para SP* and *YOLO Para SMP* architectures even further by incorporating five prediction heads. This expanded configuration enables the model to detect a full spectrum of parasite sizes, from the smallest ring-stage parasites to larger forms that may appear in advanced stages of malaria. By adding multiple prediction heads, *YOLO Para AP* provides a highly versatile detection framework capable of accurately identifying objects across a wide range of scales. This architecture is particularly valuable in clinical settings where a diverse range of parasite sizes may be present within the same blood smear, ensuring that no critical details are overlooked during the diagnostic process. The comprehensive nature of *YOLO Para AP* makes it the most robust and versatile model in the **YOLO Para** series, suitable for a wide array of diagnostic applications.

5. Experimental evaluation

In this section, we present a comprehensive evaluation of our proposed framework for the detection and classification of malaria parasites in full-size blood smear images. First, the experimental setup is

Table 4
Hyperparameters for detection models training.

Hyperparameter	Value	Description
Learning rate	0.01	Step size for gradient descent
Batch size	2	Number of samples per mini-batch
Dropout rate	0.1	Probability of dropping a neuron
Optimizer	Adam	Optimization algorithm
Weight decay	5e−4	Regularization strength
Epochs	20	Number of training epochs
Warmup epochs	3	Number of warmup epochs
Image size	1280	Image resize size

Table 5
Augmentation strategies employed on the MP-IDB [30] detection task to address data limitations, along with their associated probabilities.

Augmentation	Parameters	Probability
Rotation	Range iterations: [0, 3]	1
Gaussian Noise	Variance range: [50, 100]	0.3
HSV - Hue	Shift limit: 20	0.3
HSV - Saturation	Shift limit: 30	0.3
HSV - Value	Shift limit: 20	0.3

Table 6
Augmentation strategies used on every single dataset for the classification task, with the associated probabilities.

Augmentation	Parameters	Probability
Rotation	Range iterations: [0, 3]	1
Color jitter	Brightness: 0.2, contrast: 0.2	0.6
Gaussian blur	Kernel size: (3, 3)	0.4
Horizontal flip	–	0.5
vertical flip	–	0.5
Random crop	Crop size: 224 × 224	0.5

Table 7
Hyperparameters applied to train classification algorithms.

Hyperparameter	Value	Description
Learning rate	1e−4	Step size for gradient descent
Batch size	16	Number of samples per mini-batch
Dropout rate	0.1	Probability of dropping a neuron
Optimizer	AdamW	Optimization algorithm
Weight decay	1e−2	Regularization strength
Epochs	40	Number of training epochs

described in Section 5.1. Then, we present the experimental results, obtained on M5, IML, and MP-IDB in Sections 5.2 to 5.4, conducted to evaluate the performance of the proposed models across the following three dimensions:

(i) **Parasite detection**: to evaluate the ability to localize parasites in full-size images, as presented in Sections 5.2.1, 5.3.1 and 5.4.1 for M5, IML, and MP-IDB and its subtypes, respectively.
(ii) **Parasite stage classification**: to measure the capability to classify parasites into life stages (ring, trophozoite, schizont, gametocyte) from their single-cell image representation, as indicated in Sections 5.2.2, 5.3.2 and 5.4.2 for M5, IML and MP-IDB and its subtypes, respectively.
(iii) **Generalizability under a real-world clinical scenario**: to validate the robustness of the models across a real-world environment such as the one represented by the three different magnifications and microscope qualities provided by the M5 dataset, which exhibit distinct characteristics in Section 5.5.

5.1. Experimental setup

The framework proposed in this work and the experimental evaluations are aimed at detecting and classifying malaria parasites in full-sized images. The datasets used consist of high-resolution images of red blood cells, both infected and uninfected, covering various stages of malaria parasites, such as ring, trophozoite, schizont, and gametocyte.

Here, we detail the detection and classification setups, comprising hyperparameters and implementation choices, used for the evaluations.

Datasets splits.

To keep our code reproducible, we maintained the same splits provided by the authors of the datasets if provided or, otherwise, used the same codebases to reproduce the splitting technique used by other works. In detail, we use the same split modalities of [21,22] for the MP-IDB dataset while we follow the splits of the proposing works [46,85] for IML and M5. We employ a dropout value of 0.1% for classification and detection to mitigate overfitting in our strategies.

Detection hyperparameters and augmentations. Table 4 lists the hyperparameters for training the detection models. We employed the automatic hyperparameter selection process proposed by the YOLOv8 framework on our largest model, YOLO Para AP, and then applied the same parameters for our other models to maintain consistency across all our experiments. To enhance detection performance, as demonstrated in previous work [23], the MP-IDB was oversampled using the augmentation strategy outlined in Table 5, resulting in 35 images per original image. All experiments were conducted on this augmented dataset. Augmentations to the IML [46] and M5 [85] datasets were not required due to their already extensive image collections. Similarly to previous works [21,22,24], we employ a simpler single-class detection paradigm than a multi-class train approach. We, therefore, train our models to distinguish between parasitized versus healthy red blood cells instead of discriminating between specific malaria life stages among healthy cells

Classification hyperparameters. Table 7 shows the hyperparameters used for training the classification models. These parameters remained unchanged throughout the entire evaluation procedure to ensure fairness. We also report in Table 6 the full set of classification augmentations used during the train process.

Experiments were conducted on a workstation with an *Intel(R) Core(TM) i5-9400f* CPU running at 4.1 GHz, 32 GB RAM, and an *NVIDIA RTX 3060 GPU* with 12 GB of memory. The performance of the models was evaluated using the performance measures presented in Section 3.5.

Models selection. We have chosen specific DL architectures for the stage classification task to highlight the capabilities of the latest convolutional and transformer-based technologies against the proposal framework, either by using them for comparison purposes or as the backbone of the proposed framework.

In particular, the exploited CNN architectures were *InternImage* [86], with tiny (*internimage-t*) and small (*internimage-s*) variants, and *ConvNextV2* [87], which includes tiny (*convnextv2-t*) and base (*convnextv2-b*) models. Instead, the following ViT-based architectures were used: *Dino* [88], with small (*dino-s*) and base (*dino-b*) versions; *SwinV2* [110], featuring tiny (*swinv2-t*) and base (*swinv2-b*) configurations; and *ViT* [89], which includes base (*vit-b*) and large (*vit-l*) models.

5.2. Results on M5

This subsection presents the results obtained on the M5 [85].

5.2.1. Detection results

As presented in Table 8 the produced results showcase the performances of the different YOLO Para models in comparison to the baseline ones. The best results in terms of $AP_{0.50:0.95}$ were reached by the YOLO Para SMP model and YOLO Para SP model, both reaching an $AP_{0.50:0.95}$ score of 70%.

Table 8
Experimental results obtained on the M5 dataset [85] The reported performance metrics include Average Precision at different Intersection over Union thresholds. The number of parameters for each model is also provided.

Model	$AP_{0.50:0.95}$	AP_{50}	AP_S	AP_M	AP_L	Params (M)
YOLO PAM [22]	0.70	0.94	–	–	0.70	48
YOLOv8m	0.69	0.95	–	–	0.69	34
YOLOv8l	0.68	0.93	–	–	0.68	77
YOLO Para SP	0.70	0.95	–	–	0.70	39
YOLO Para SMP	**0.71**	**0.96**	–	–	**0.71**	51
YOLO Para AP	0.68	0.94	–	–	0.68	68
YOLO SPAM++ [21]	0.70	0.95	–	–	0.70	30
YOLO SPAM [21]	0.67	0.93	–	–	0.67	**23**
YOLOv5m	0.69	0.94	–	–	0.69	25
YOLOv5l	0.68	0.93	–	–	0.68	53

Table 9
Experimental results on the M5 dataset [85] for the Falciparum stage classification task. The models were trained on the original crops from the training set, and evaluated on the detected crops from the test set.

Model	ACC	Macro F1	Macro PRE	Macro REC	Params (M)
internimage-t	**0.91**	0.58	0.59	0.58	29
internimage-s	0.72	0.34	0.36	0.33	50
dino-s	0.50	0.34	0.38	0.40	**21**
dino-b	0.90	0.53	0.53	0.52	85
convnextv2-t	0.89	0.50	0.50	0.51	27
convnextv2-b	0.89	0.52	0.59	0.49	87
swinv2-t	0.88	0.54	0.53	0.55	27
swinv2-b	**0.91**	**0.67**	**0.88**	0.61	86
vit-b	**0.91**	0.65	0.69	**0.62**	86
vit-l	0.90	0.54	0.55	0.54	303

Table 10
Experimental results obtained on the IML dataset [46]. The reported performance metrics include Average Precision at different IoU thresholds. The number of parameters for each model is also provided.

Model	$AP_{0.50:0.95}$	AP_{50}	AP_S	AP_M	AP_L	Params (M)
YOLO PAM [22]	0.599	0.918	–	0.600	0.650	48
YOLOv8m	0.562	0.892	–	0.555	0.640	34
YOLOv8l	0.571	0.901	–	0.565	0.655	77
YOLO Para SP	0.597	0.911	–	0.590	0.655	39
YOLO Para SMP	**0.674**	**0.944**	–	**0.670**	**0.710**	51
YOLO Para AP	0.626	0.908	–	0.620	0.675	68
YOLO SPAM++ [21]	0.615	0.869	–	0.610	0.640	30
YOLO SPAM [21]	0.62	0.891	–	0.615	0.640	**23**
YOLOv5m	0.604	0.865	–	0.600	0.625	25
YOLOv5l	0.591	0.83	–	0.585	0.605	53

Table 11
Experimental results on the IML dataset [46] for the Vivax stage classification task. The models were trained on the original crops from the training set and evaluated on the detected crops from the test set.

Model	ACC	Macro F1	Macro PRE	Macro REC	Params (M)
internimage-t	0.653	0.561	0.558	0.616	29
internimage-s	0.295	0.224	0.246	0.224	50
dino-s	0.537	0.416	0.442	0.413	**21**
dino-b	**0.758**	0.57	0.564	0.577	85
convnextv2-t	0.705	0.536	0.556	0.543	27
convnextv2-b	0.747	0.561	0.555	0.57	87
swinv2-t	0.737	0.555	0.55	0.562	27
swinv2-b	0.747	0.563	0.589	0.577	86
vit-b	**0.758**	**0.727**	**0.813**	**0.688**	86
vit-l	0.747	0.562	0.557	0.572	303

Table 12
Experimental results obtained on the MP-IDB dataset [30] on the Falciparum class. The reported performance metrics include Average Precision at different IoU thresholds. The number of parameters for each model is also provided.

Model	$AP_{0.50:0.95}$	AP_{50}	AP_S	AP_M	AP_L	Params (M)
YOLO PAM [22]	0.836	0.989	0.76	0.8	**1**	48
YOLOv8m	0.789	0.983	0.7	0.77	0.9	34
YOLOv8l	0.809	0.985	0.73	0.79	0.9	77
YOLO Para SP	0.858	**0.991**	0.78	**0.83**	0.8	39
YOLO Para SMP	0.857	0.991	0.78	0.82	**1**	51
YOLO Para AP	**0.865**	**0.991**	**0.79**	**0.83**	**1**	68
YOLO SPAM++ [21]	0.846	**0.991**	0.755	0.803	**1**	30
YOLO SPAM [21]	0.747	0.987	0.656	0.698	**1**	**23**
YOLOv5m	0.811	0.985	0.72	0.79	0.9	25
YOLOv5l	0.782	0.982	0.69	0.77	0.9	53

Table 13
Experimental results obtained on the MP-IDB dataset [30] on the Malariae class. The reported performance metrics include Average Precision at different IoU thresholds. The number of parameters for each model is also provided.

Model	$AP_{0.50:0.95}$	AP_{50}	AP_S	AP_M	AP_L	Params (M)
YOLO PAM [22]	0.788	0.972	0.34	0.74	–	48
YOLOv8m	0.929	0.994	0.9	0.84	–	34
YOLOv8l	0.913	0.993	0.8	0.84	–	77
YOLO Para SP	0.95	**0.995**	0.83	0.85	–	39
YOLO Para SMP	0.946	**0.995**	**0.93**	0.84	–	51
YOLO Para AP	**0.949**	**0.995**	0.8	0.84	–	68
YOLO SPAM++ [21]	0.936	0.985	0.85	0.842	–	30
YOLO SPAM [21]	0.941	**0.995**	0.88	**0.859**	–	**23**
YOLOv5m	0.872	0.98	0.75	0.78	–	25
YOLOv5l	0.897	**0.995**	0.85	0.83	–	53

5.2.2. Classification results

The classification results for the stage classification task on the M5 dataset are depicted in Table 9. The best results are achieved using the Swinv2 base model, reaching an F1 score of 67%.

5.3. Results on IML

This subsection gives the results obtained on the IML [46].

5.3.1. Detection results

As presented in Table 10 the produced results showcase the performances of the different YOLO Para models in comparison to the baseline ones. The best results in terms of $AP_{0.50:0.95}$ were reached by the YOLO Para SMP model, reaching an $AP_{0.50:0.95}$ score of 67.4%.

5.3.2. Classification results

The classification results for the stage classification task on the IML dataset are depicted in Table 11, the best results are achieved using Vit base, reaching a F1 score of 72.7%.

5.4. Results on MP-IDB

This subsection provides a comprehensive set of tables that display the results obtained from the MP-IDB dataset [30].

5.4.1. Detection results

As depicted in Tables 12 to 15 the best detection results for the MP-IDB dataset over the different species are reached using YOLO Para SMP and YOLO Para AP models. The first one reaches the new state-of-the-art results for the P. Ovale class while the YOLO Para AP reaches also state-of-the-art results for the P. Falciparum, P. Malariae, and P. Vivax classes.

5.4.2. Classification results

The classification results for the stage classification task on the MP-IDB dataset P. Falciparum class are depicted in Table 16, the best results are achieved using a ViT base reaching an F1 of 79%.

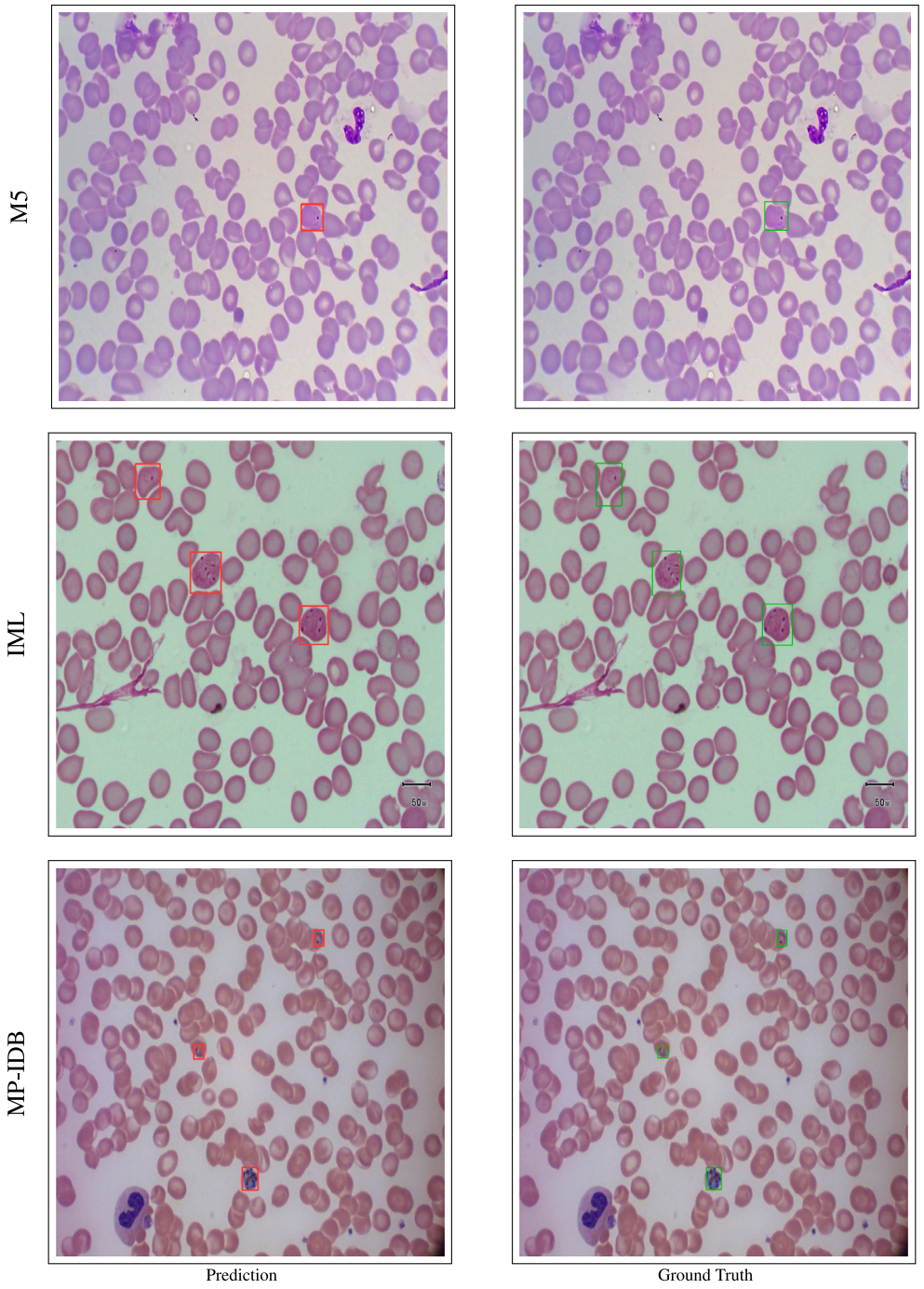


Fig. 6. Comparison of predictions from the trained YOLO Para AP architecture with ground truth annotations. The left column shows the model's predictions, which are highlighted with green bounding boxes, while the right column displays the ground truth, with red bounding boxes marking the malaria parasites. The first row presents images from the MP-IDB dataset, the second row from the IML dataset, and the third row from the M5 dataset.

5.5. Generalizability analysis of YOLO Para under a real-world clinical scenario

In this section, we present our efforts toward developing improved architectures for malaria identification. Our evaluation includes not only performance measures obtained by training and testing on the same source but also an analysis of performance correlations across different microscope types and magnification levels. To effectively assess our model's capabilities in such scenarios, we utilize the M5 dataset [85], which uniquely includes various modalities absent in other datasets.

We report our findings in Table 17, where the tests were conducted on both the YOLO Para SMP model and the YOLO model used in the [85] evaluation for a fair comparison. Our model demonstrates robust performance across diverse evaluation datasets, showing that it can generalize well to variations in microscope quality without requiring domain generalization techniques. Furthermore, we observe that performance decreases slightly under substantial magnification changes, such as from 1000× to 400×, with a 6.88% drop from 400× to 1000× and a 10.65% drop from 1000× to 400×.

Table 14
Experimental results obtained on the MP-IDB dataset [30] on the Ovale class. The reported performance metrics include Average Precision at different IoU thresholds. The number of parameters for each model is also provided.

Model	$AP_{0.50:0.95}$	AP_{50}	AP_S	AP_M	AP_L	Params (M)
YOLO PAM [22]	0.944	**0.995**	–	0.85	–	48
YOLOv8m	0.897	**0.995**	–	0.83	–	34
YOLOv8l	0.776	0.993	–	0.72	–	77
YOLO Para SP	0.932	**0.995**	–	0.83	–	39
YOLO Para SMP	**0.951**	**0.995**	–	**0.86**	–	51
YOLO Para AP	0.935	**0.995**	–	0.85	–	68
YOLO SPAM++ [21]	0.874	0.928	–	0.792	–	30
YOLO SPAM [21]	0.938	**0.995**	–	0.839	–	**23**
YOLOv5m	0.902	**0.995**	–	0.83	–	25
YOLOv5l	0.9	**0.995**	–	0.81	–	53

Table 15
Experimental results obtained on the MP-IDB dataset [30] on the Vivax class. The reported performance metrics include Average Precision at different IoU thresholds. The number of parameters for each model is also provided.

Model	$AP_{0.50:0.95}$	AP_{50}	AP_S	AP_M	AP_L	Params (M)
YOLO PAM [22]	0.872	0.942	0.19	**0.86**	0.94	48
YOLOv8m	0.859	0.937	0.19	0.83	**0.96**	34
YOLOv8l	0.836	0.936	0.17	0.835	0.930	77
YOLO Para SP	0.866	0.94	**0.22**	0.83	0.95	39
YOLO Para SMP	0.875	0.944	0.16	0.83	0.91	51
YOLO Para AP	**0.883**	**0.946**	0.19	**0.86**	0.95	68
YOLO SPAM++ [21]	0.875	0.935	0.152	0.81	0.89	30
YOLO SPAM [21]	0.836	0.929	0.152	0.792	0.898	**23**
YOLOv5m	0.826	0.935	0.16	0.83	0.95	25
YOLOv5l	0.831	0.932	0.156	0.811	0.924	53

Table 16
Experimental results on the MP-IDB dataset [30] for the Falciparum stage classification task. The models were trained on the original crops from the training set and evaluated on the detected crops from the test set.

Model	ACC	Macro F1	Macro PRE	Macro REC	Params (M)
internimage-t	0.794	0.564	0.532	0.777	29
internimage-s	0.886	0.627	0.615	0.682	50
dino-s	**0.938**	0.649	0.796	0.579	**21**
dino-b	0.931	0.747	0.705	0.831	85
convnextv2-t	0.858	0.582	0.577	0.769	27
convnextv2-b	0.909	0.719	0.704	0.784	87
swinv2-t	0.681	0.681	0.742	0.802	27
swinv2-b	0.58	0.618	0.742	0.731	86
vit-b	0.901	**0.79**	**0.798**	**0.85**	86
vit-l	0.912	0.731	0.818	0.734	303

Table 17
$AP_{0.50:0.95}$ values for different train and test dataset combinations, grouped by testing dataset using the M5 [85] dataset.

Model	Train dataset	$AP_{0.50:0.95}$
Test dataset: 100 × HCM		
YOLO Para SMP	100 × HCM	0.17
YOLO [85]	100 × HCM	0.20
YOLO Para SMP	400 × HCM	0.05
YOLO [85]	400 × HCM	0.05
YOLO Para SMP	1000 × HCM	0.003
YOLO [85]	1000 × HCM	0.00
YOLO Para SMP	100 × LCM	0.09
YOLO Para SMP	400 × LCM	0.03
YOLO Para SMP	1000 × LCM	0.04
Test dataset: 400 × HCM		
YOLO Para SMP	100 × HCM	0.06
YOLO [85]	100 × HCM	0.04
YOLO Para SMP	400 × HCM	0.60
YOLO [85]	400 × HCM	0.57
YOLO Para SMP	1000 × HCM	0.53
YOLO [85]	1000 × HCM	0.37
YOLO Para SMP	100 × LCM	0.08
YOLO Para SMP	400 × LCM	0.35
YOLO Para SMP	1000 × LCM	0.32
Test dataset: 1000 × HCM		
YOLO Para SMP	100 × HCM	0.09
YOLO [85]	100 × HCM	0.11
YOLO Para SMP	400 × HCM	0.60
YOLO [85]	400 × HCM	0.55
YOLO Para SMP	1000 × HCM	0.71
YOLO [85]	1000 × HCM	0.63
YOLO Para SMP	100 × LCM	0.07
YOLO Para SMP	400 × LCM	0.29
YOLO Para SMP	1000 × LCM	0.32
Test dataset: 100 × LCM		
YOLO Para SMP	100 × HCM	0.03
YOLO Para SMP	400 × HCM	0.01
YOLO Para SMP	1000 × HCM	0.0002
YOLO Para SMP	100 × LCM	0.09
YOLO Para SMP	400 × LCM	0.02
YOLO Para SMP	1000 × LCM	0.01
Test dataset: 400 × LCM		
YOLO Para SMP	100 × HCM	0.08
YOLO Para SMP	400 × HCM	0.23
YOLO Para SMP	1000 × HCM	0.06
YOLO Para SMP	100 × LCM	0.13
YOLO Para SMP	400 × LCM	0.35
YOLO Para SMP	1000 × LCM	0.20
Test dataset: 1000 × LCM		
YOLO Para SMP	100 × HCM	0.12
YOLO Para SMP	400 × HCM	0.27
YOLO Para SMP	1000 × HCM	0.13
YOLO Para SMP	100 × LCM	0.10
YOLO Para SMP	400 × LCM	0.34
YOLO Para SMP	1000 × LCM	0.46

6. Discussion

This section discusses the results of the experimental evaluation conducted. We start in Section 6.1 with the ablation study conducted to better motivate the architectural design used in this study. Then, the detection and classification experiment results are given in Sections 6.2 and 6.3. In addition, further analyses are provided in Sections 6.4 to 6.6, based on the parasite size, computational performance, and low parasitemia results. Finally, in Section 6.7, we compare the obtained results with those available from the literature work and give the final insights from the experimentations in Section 6.8.

6.1. Ablation study on YOLO Para SMP architectural choices

To guide the architectural design of the YOLO Para architectures, we conducted an ablation study evaluating the impact of different attention mechanisms on performance. Specifically, we tested configurations without CBAM, NAM and without C3STR blocks to assess how these components contribute to the model's robustness and accuracy on the M5 dataset [85] using the YOLO Para SMP as a guideline model. By isolating these factors, we aimed to identify the configuration that best supports generalization across diverse microscope types and magnification levels.

As shown in Table 18, removing CBAM or NAM individually leads to slight performance declines, while removing both results in the largest drop in $AP_{0.50:0.95}$ score. These results suggest that both attention mechanisms contribute positively to the model's performance and that their combined use offers the best results for the architecture, the C3STR module makes the biggest contribution in terms of performance metrics and combined with the other attention modules provides the best overall score.

Table 18
Performance of YOLO Para SMP on the M5 dataset [85] with different attention strategies.

Attention strategy	AP
YOLOv8m	0.690
YOLOv8m+CBAM	0.699
YOLOv8m+NAM	0.696
YOLOv8m+C3STR	0.705
YOLOv8m+NAM+CBAM	0.702
YOLO Para SMP	0.711

Table 19
Summary of off-the-shelf and proposed YOLO models with parameters (in millions), GFLOPs, and inference speeds. Values evaluated on RTX 3060.

Model	Parameters (M)	GFLOPs	Inference speed (ms)
YOLO Para SP	39	139.2	40.6
YOLO Para SMP	51	143.3	43.5
YOLO Para AP	68	163.4	73.1
YOLOv8m	34	86.0	25.9
YOLOv8l	77	192.8	45.9
YOLOv5m	25	64.6	25.6
YOLOv5l	53	135.6	34.0

6.2. Detection results overview

As observable from Tables 8, 10 and 12 to 15 across all the different detection experiments, one pattern emerges.

The models proposed in each experiment outperform the baseline YOLOv5 and YOLOv8 models. Specifically, the YOLO Para AP and YOLO Para SMP models tend to perform better across the novelties. This is likely due to the fact that the majority of parasites in the datasets are medium and small in size, which both architectures excel at detecting. The YOLO Para SMP model is the best one for the M5 dataset, although the YOLO PAM [22] model comes in second.

In addition, we report a visual representation of the prediction capabilities of the proposed YOLO Para AParchitecture across the used datasets in Fig. 6

6.3. Stage classification results overview

In three classification experiments depicted in Tables 9, 11 and 16 the results show a clear pattern of better performances reached by using transformer-based models. In particular, for larger datasets such as M5 the Swinv2 transformer provided the best results while the plain Vit base model outperformed all other methods in the other smaller datasets.

6.4. Parasite size quantification

The subsequent analysis is devoted to diagnosing and quantifying the dimensions of the detected parasites to better understand the results obtained from the experimentations realized.

The results reported use common evaluation metrics at different object sizes, such as AP_L, AP_S, and AP_M, which are often considered the standard for most evaluation protocols. However, these kinds of metrics are insufficient in the case of parasites, which are often represented by various shapes. In our previous work, a deep learning framework for malaria detection focused on the issue of P. Falciparum detection [23] in MP-IDB. Despite the fact that it produced state-of-the-art results, it presented limitations in tiny ring stage parasite detection. Specifically, it shows how one of the employed models (i.e.,YOLO SPAM++ [21]) addresses the issues, reaching outstanding results.

The study of size quantification is important to define an appropriate concept of size and show the inherent differences within the datasets. Different datasets can be obtained with varying types of equipment, which often produce different brightness variations, but, more importantly, different magnifications. To conduct this experiment, the most effective approach is to calculate a histogram of the different areas grown in the datasets. As shown by the results in Fig. 7, MP-IDB presents tiny parasites compared to the others. In contrast, IML offers the most significant area-related parasites, and, finally, M5 has the most balanced distribution.

After these considerations and results, it is possible to define a new definition of small, medium, and large-sized parasites. Fig. 7(a) shows that the areas can be empirically subdivided into three different thresholds:

- less than $0.2 * 1e{-}2$: **small**-sized parasites;
- greater than $0.6 * 1e{-}2$: **large**-sized parasites;
- otherwise: **medium**-sized parasites.

Further analysis was conducted on the best model for each dataset using these new discrimination criteria, thus modifying the standard COCO evaluation. According to Fig. 8, the overall best-performing models for tiny parasites are YOLO Para SP and YOLO Para AP. These results can be justified by the addition, for both, of a prediction head dedicated to the first backbone feature map, excluding the first channel expansion convolution. There is no clear indication of which models for other parasite sizes. However, on average, YOLO Para SMP obtained excellent performance in almost all cases of study and is in line with the best when it is not. As mentioned in the description of Fig. 4, this model uses three prediction heads, none specialized for extremely small or large objects. These results are predictable because each proposed model considers intermediate feature maps suitable for medium and large objects (see Figs. 9 and 10).

6.5. Computational analysis

Despite improvements in reference metrics, our models that incorporate Swin Transformer-based modules experience increased GFLOPs, as quantified in Table 19, which also affects inference speed. However, larger off-the-shelf models like YOLOv5 and YOLOv8 large variants demonstrate comparable performance, making them viable for near real-time applications.

6.6. Low parasitemia results discussion

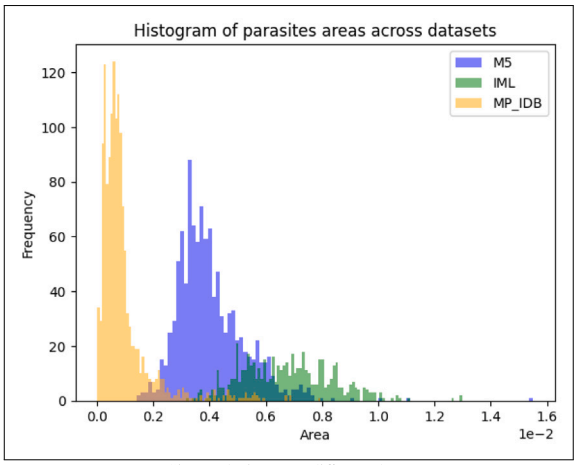
Our proposed architectures not only outperform previous ad-hoc models in detecting small parasites but also demonstrate remarkable robustness in recognizing parasites under low parasitemia conditions. This includes challenging classes such as P. Malariae and P. Ovale in the MP-IDB dataset [30], where the parasite-to-image ratios are as low as 1.16 and 1.13, respectively. This enhancement leads to an improvement over previous state-of-the-art performance with an increase in Average Precision by 0.8% and 0.7% for these classes. Key strategies implemented to achieve these gains include tailored data augmentation techniques to enhance detection in low-parasitemia cases, which proved to be pivotal in the development process as described in Table 4

6.7. Comparison with state of the art work

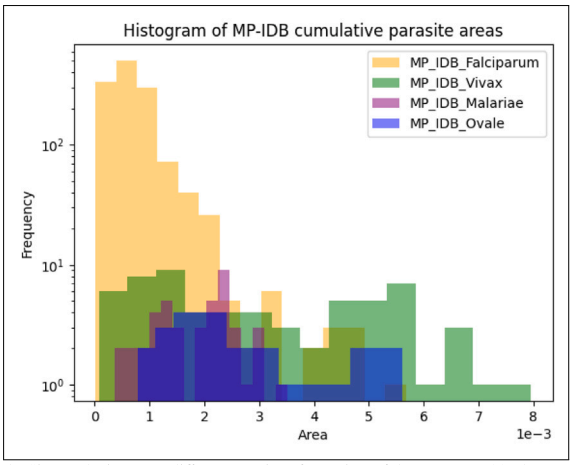
A comparison of the previous state of the art and our proposed framework is given in Tables 20 and 21 for the detection and classification tasks, respectively.

In Table 20, a significant improvement in terms of $AP_{0.50:0.95}$ is observed in every single dataset and also in each subset of MP-IDB. These results clearly show the superiority of the proposed framework on the detection task.

As regards the classification performance, there are some important observations to note. First, the approaches presented in Table 21 do not follow the same evaluation procedure. In particular, Zedda et al. [24]



(a) Size analysis across different datasets.



(b) Size analysis across different species of parasites of the MP-IDB [30] dataset.

Fig. 7. Comparison of the parasites' sizes of the different adopted datasets [30,46,85].

Table 20
Comparison of our proposed framework against the state-of-the-art in terms of detection performance.

Dataset	Species	Work	Reference model	AP	Increase AP%
M5	P. Falciparum	Sultani et al. (2022) [85]	Faster R-CNN	66.8	–
M5	P. Falciparum	Proposed framework	YOLO Para SMP	**71.0**	**4.2**
MP-IDB	P. Falciparum	Zedda et al. [21]	YOLO SPAM++	84.6	–
MP-IDB	P. Falciparum	Proposed framework	YOLO Para AP	**86.5**	**1.9**
MP-IDB	P. Malariae	Zedda et al. [21]	YOLO SPAM	94.1	–
MP-IDB	P. Malariae	Proposed framework	YOLO Para $AP_{0.50:0.95}$	**94.9**	**0.8**
MP-IDB	P. Ovale	Zedda et al. [22]	YOLO PAM	94.4	–
MP-IDB	P. Ovale	Proposed framework	YOLO Para SMP	**95.1**	**0.7**
MP-IDB	P. Vivax	Zedda et al. [21]	YOLO SPAM++	87.5	–
MP-IDB	P. Vivax	Proposed framework	YOLO Para AP	**88.3**	**0.8**
IML	P. Vivax	Proposed framework	YOLO Pra SMP	**67.4**	–

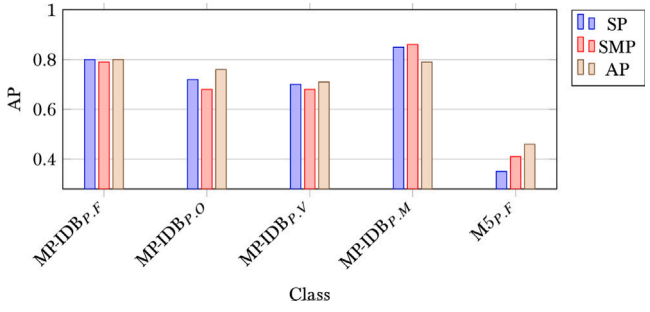


Fig. 8. Average precision of small objects detection performances for each YOLO SPAM model, in the figure are depicted the performances for MP-IDB P. Falciparum, P. Ovale, P. Vivax, P. Malariae and M5 P. Falciparum.

and our approach are the only ones that perform the life-stage classification on the parasites detected by the detection framework rather than on a test set sampled by the entire crops available. This was done to ensure an overall quantification of the performance of the entire framework by simulating a more realistic clinical scenario, where the parasite crops are effectively provided after a first stage of analysis from full-size images.

Although our classification method is not reported as the best-performing in Table 21, we considered the results of 93.80% in terms of accuracy highly satisfactory because of the previous considerations.

6.8. Significance and implications of the results

The experimental results demonstrate the proposed framework's ability to address key challenges in malaria diagnosis:

- **Enhanced detection**: YOLO Para SMP achieved up to a 23% improvement in $AP_{0.50:0.95}$ over baseline methods for detecting small parasites.
- **Stage classification**: the models accurately distinguished parasite life stages, particularly on imbalanced datasets, through data augmentation and attention-based architectures.
- **Real-world adaptation** the proposed framework is well-suited for deployment in real-world clinical and low-resource settings by addressing the typical variations encountered in the microscope quality and magnification scenario provided by the M5 dataset.
- **Comparison with traditional microscopy** Traditional microscopy is often the golden standard for malaria identification, however, such technique is prone to human error and is highly influenced by the experience of the medical personnel, in [112] is it shown that the accuracy for the studied cases is 91%. Notably, our models for the IML and MP-IDB reach near-perfect results in terms of AP_{50} not only emphasizing high accuracy but also precision in the detected bounding boxes.

6.9. Limitations

Limitations from a clinical perspective. The proposed YOLO Para model series has demonstrated promising results in accurately identifying malaria parasites from Giemsa-stained slides. However, the clinical context in which these models may be deployed poses some inherent

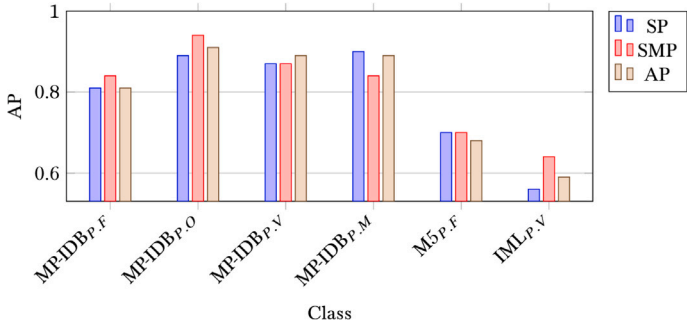


Fig. 9. Average precision of medium sized objects detection performances for each YOLO SPAM model, in the figure are depicted the performances for MP-IDB P. Falciparum, P. Ovale, P. Vivax, P. Malariae, M5 P. Falciparum and IML P. Vivax parasites.

Table 21
Comparison of performance with the state of the art on the classification task.

Work	Task	Method	ACC (%)
Rahman et al. [33]	Single cells stages classification	VGG-19	85.18
Maity et al. [8]	Segmentation+classification of ring stage	ANN + CapsNet	95.46
Zedda et al. [24]	Detection+classification of all stages	YOLOv5 + DarkNet-53	96.05
Chen et al. [111]	Segmentation+classification of types+classification of all stages	U-Net + SDU-Net + MobileNetV1	98.83
Proposed approach	Detection+classification of all stages	YOLO Para AP + dino-s	93.80

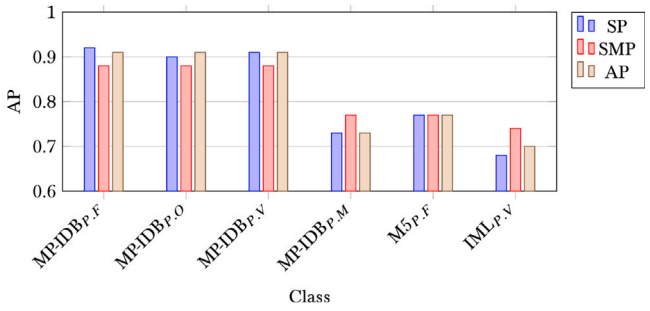


Fig. 10. Average precision of large-sized objects detection performances for each YOLO SPAM model, in the figure are depicted the performances for MP-IDB P. Falciparum, P. Ovale, P. Vivax, P. Malariae, M5 P. Falciparum and IML P. Vivax parasites.

challenges.

One significant limitation of our work is the potential vulnerability to false positives, particularly in distinguishing malaria from other diseases that exhibit similar morphological features in blood smears. For instance, other plasmodial infections or conditions such as leukemias can present cells that may be misclassified as malaria parasites (e.g.,altered red blood cells, irregular leukocytes).

To mitigate the risk of misdiagnosis in clinical settings with the proposed approach, we propose some guidelines that can be pursued in future developments of the current work: (i) future development of the model could benefit from the inclusion of a broader range of data that encompasses various diseases characterized by similar appearances to enhance the model's ability to discern malaria parasites amidst other objects; (ii) implementing a multi-class classification system that does not solely focus on malaria detection could improve the model's specificity while learning to differentiate not just the target species but also various aberrant blood cell presentations indicative of other diseases; (iii) before widespread deployment of our model in clinical practice, rigorous validation against comprehensive control datasets should be conducted. Collaborations with clinical experts in hematology may provide more robust training and validation phases, ensuring that the model can accurately discern malaria from other conditions; (iv) when incorporated into applications or software utilized in clinical settings, including contextual clinical guidelines and decision-support tools, can be advantageous. This would enable practitioners to interpret results in conjunction with traditional diagnostics, thereby reducing the risks associated with potential misclassification.

Limitations from the datasets perspective. In this study, we introduced a deep learning framework for the detection and classification of malaria parasites using Giemsa-stained blood smear images from three publicly available datasets. While our model demonstrates significant improvements in detection accuracy, the quality of the input images considerably impacts the model's performance.

In particular, the public datasets used in this research exhibit various realistic imaging challenges, including but not limited to inconsistent background illumination, variations in stain quality, and differences in imaging conditions (e.g., magnification and microscope type). To maximize the performance of our model series, we provide some guidelines to follow: (i) for improved detection accuracy, it is recommended that images be captured against a consistent background color. A white or neutral background minimizes visual distractions and enhances the model's ability to focus on the features of interest. This is particularly important when the images originate from varied sources where background consistency may not be guaranteed; (ii) ensuring proper white balance during image acquisition is also important. Images that are too bright or too dark can compromise the visibility of malaria parasites. Implementing preprocessing steps to standardize brightness and contrast across images can help enhance feature extraction by the model; (iii) rigorous quality control measures should be implemented to assess whether images meet the established criteria for inclusivity within the training and testing datasets. This may involve filtering out images with notable artifacts or misclassifications before model training; (iv) it is mandatory to develop a standard operating procedure for image capture that outlines best practices, including optimal lighting conditions, camera settings, and image formats. Clear guidance on minimizing background noise can facilitate the collection of high-quality images that enhance the robustness of the model.

Future work will include extensive validation under varied imaging conditions to further refine these guidelines and optimize model performance.

7. Conclusion

This study presents a novel computer vision-based pipeline for malaria parasite detection, integrating attention mechanisms and the YOLO architecture to achieve superior detection and classification of

malaria parasites across various life stages and species. The experimental evaluation demonstrates that the proposed YOLO Para series, particularly the YOLO Para SMP and YOLO Para AP models, consistently outperform baseline approaches in terms of precision, recall, and F1 scores on three benchmark datasets (M5, IML, and MP-IDB). These results highlight the robustness of the proposed architecture, especially in detecting small-scale ring-stage parasites critical for early diagnosis. The performance gains observed are attributed to key architectural innovations, such as the use of attention mechanisms to enhance feature extraction, multi-head configurations for scale-aware detection, and augmentation strategies for handling class imbalance. In particular, the YOLO Para AP model exhibited a remarkable ability to generalize across multiple datasets with diverse imaging characteristics, emphasizing its potential for real-world applications in resource-constrained environments.

Despite our models' small computational trade-off, we show that they also improve upon the performances in the malaria detection field without sacrificing the near-real-time capabilities of off-the-shelf models. They also showed strong capabilities across different scenarios, such as low-parasitemia for the MP-IDB's P. Malariae and P. Ovale classes and, most importantly, strong generalization capabilities across microscope quality and magnification, as experimented on the M5 dataset.

The achievements presented in this study have also revealed avenues for further research. Challenges such as handling extremely low parasitemia levels, detecting mixed infections, and adapting models for low-cost mobile imaging devices remain largely unexplored. Future directions could include: (1) refining existing architectures or developing new transformer-based models to enhance accuracy and efficiency. Most importantly, the focus will be on developing more lightweight architectures suitable for fully real-time scenarios without sacrificing the near-optimal performance achieved by our YOLO Para architectures. A key aspect of this development could be the integration of novel fast attention mechanisms, such as Flash Attention [113,114]; (2) obtain valuable insights into their generalization and robustness by testing these models on benchmark datasets like COCO with possible investigations on domain adaptation techniques to improve model robustness across varying imaging conditions and devices. Examples include the incorporation of contrastive pretraining [82] into our downstream pipelines, facilitating smoother adaptation for detection even in new domains and the expansion of datasets to include additional Plasmodium species and diverse geographic variations; (3) exploring the transferability of these models to related tasks, such as detecting other types of parasites or objects in medical images, is also a key area for further research. This could involve fine-tuning new datasets or designing specialized models, for example, investigating further improvements of transformer-based architectures to capture long-range dependencies for more precise classification of parasite stages; (4) improving interpretability and explainability by visualizing learned features to enhance the understanding of model predictions.

Acronyms

See Table 22.

Table 22
List of acronyms used in this paper.

Acronym	Meaning
DL	Deep Learning
CNN	Convolutional Neural Network
YOLO	You Only Look Once
CAD	Computer-Aided Detection
CNN	Convolutional Neural Network
SVM	Support Vector Machine
RDT	Rapid Diagnostic Test
PCR	Polymerase Chain Reaction
WHO	World Health Organization
HBHI	High Burden to High Impact
ViT	Vision Transformer
CBAM	Convolutional Block Attention Module
NAM	Normalized Attention Module
SPP	Spatial Pyramid Pooling
PANet	Path Aggregation Network
HCM	High-Cost Microscope
LCM	Low-Cost Microscope
GPU	Graphics Processing Unit
GFLOPs	Giga Floating Point Operations

(*continued on next page*)

Table 22 (*continued*).

CRediT authorship contribution statement

Luca Zedda: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrea Loddo:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Cecilia Di Ruberto:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Code availability

The code for this study is available at the following GitHub repository: https://github.com/Snarci/YOLO-Para.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 - Call for tender No. 3277 published on December 30, 2021 by the Italian Ministry of University and Research (MUR) funded by the European Union – NextGenerationEU. Project Code ECS0000038 – Project Title eINS Ecosystem of Innovation for Next Generation Sardinia – CUP F53C22000430001- Grant Assignment Decree No. 1056 adopted on June 23, 2022 by the Italian Ministry of University and Research (MUR).

References

[1] W.H. Organization, Report of the First and Second Meetings of the Technical Advisory Group on Malaria Elimination and Certification, 13–14 September 2022 and 27 January 2023, World Health Organization, 2023.
[2] W.H. Organization, 2023, https://www.who.int/news-room/fact-sheets/detail/malaria, Online; (accessed 29 May 2023).
[3] S. Healthcare, 2021, https://stanfordhealthcare.org/medical-conditions/primary-care/malaria/types.html, Online; (accessed 29 May 2023).
[4] U.S.C. for Disease Control, Prevention, 2021, https://www.cdc.gov/malaria/about/biology/index.html, Online; (accessed 29 May 2023).
[5] A.M. Gimenez, R.F. Marques, M. Regiart, D.Y. Bargieri, Diagnostic methods for non-falciparum malaria, Front. Cell. Infect. Microbiol. 11 (2021) 681063.
[6] A. Vijayalakshmi, B.R. Kanna, Deep learning approach to detect malaria from microscopic images, Multimedia Tools Appl. 79 (21–22) (2020) 15297–15317.
[7] A. Loddo, C.D. Ruberto, M. Kocher, Recent advances of malaria parasites detection systems based on mathematical morphology, Sensors 18 (2) (2018) 513.
[8] M. Maity, A. Jaiswal, K. Gantait, J. Chatterjee, A. Mukherjee, Quantification of malaria parasitaemia using trainable semantic segmentation and capsnet, Pattern Recognit. Lett. 138 (2020) 88–94.

[9] P. Berzosa, A. de Lucio, M. Romay-Barja, Z. Herrador, V. González, L. García, A. Fernández-Martínez, M. Santana-Morales, P. Ncogo, B. Valladares, et al., Comparison of three diagnostic methods (microscopy, RDT, and PCR) for the detection of malaria parasites in representative samples from Equatorial Guinea, Malar. J. 17 (1) (2018) 1–12.

[10] A. Onken, C.G. Haanshuus, M.K. Miraji, M. Marijani, K.O. Kibwana, K.A. Abeid, K. Mørch, M. Reimers, N. Langeland, F. Müller, et al., Malaria prevalence and performance of diagnostic tests among patients hospitalized with acute undifferentiated fever in Zanzibar, Malar. J. 21 (1) (2022) 1–8.

[11] Z. Liang, A. Powell, I. Ersoy, M. Poostchi, K. Silamut, K. Palaniappan, P. Guo, M.A. Hossain, S.K. Antani, R.J. Maude, J.X. Huang, S. Jaeger, G.R. Thoma, CNN-based image analysis for malaria diagnosis, in: IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2016, Shenzhen, China, December 15-18, 2016, IEEE Computer Society, 2016, pp. 493–496.

[12] S. Rajaraman, S.K. Antani, M. Poostchi, K. Silamut, M.A. Hossain, R.J. Maude, S. Jaeger, G.R. Thoma, Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images, PeerJ 6 (2018) e4568.

[13] S. Rajaraman, S. Jaeger, S.K. Antani, Perf. eval. of deep neural ensembles toward malaria parasite detection in thin-blood smear images, PeerJ 7 (2019) e6977.

[14] C. Di Ruberto, A. Dempster, S. Khan, B. Jarra, Analysis of infected blood cell images using morphological operators, Image Vis. Comput. 20 (2) (2002) 133–146.

[15] F.B. Tek, A.G. Dempster, I. Kale, Malaria parasite detection in peripheral blood images, in: M.J. Chantler, R.B. Fisher, E. Trucco (Eds.), Proceedings of the British Machine Vision Conference 2006, Edinburgh, UK, September 4-7, 2006, British Machine Vision Association, 2006, pp. 347–356.

[16] S.K. Kumarasamy, S. Ong, K.S. Tan, Robust contour reconstruction of red blood cells and parasites in the automated identification of the stages of malarial infection, Mach. Vis. Appl. 22 (3) (2011) 461–469.

[17] S. Bias, S. Reni, I. Kale, Mobile hardware based implementation of a novel, efficient, fuzzy logic inspired edge detection technique for analysis of malaria infected microscopic thin blood images, in: -, in: Procedia Computer Science, vol. 141, Elsevier, 2018, pp. 374–381.

[18] A. Loddo, L. Putzu, On the effectiveness of leukocytes classification methods in a real application scenario, AI 2 (3) (2021) 394–412.

[19] M. Zaid, S. Ali, M. Ali, S. Hussein, A. Saadia, W. Sultani, Identifying out of distribution samples for skin cancer and malaria images, Biomed. Signal Process. Control 78 (2022) 103882.

[20] W. Sultani, W. Nawaz, S. Javed, M.S. Danish, A. Saadia, M. Ali, Towards low-cost and efficient malaria detection, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, IEEE, 2022, pp. 20655–20664.

[21] L. Zedda, A. Loddo, C. Di Ruberto, A deep architecture based on attention mechanisms for effective end-to-end detection of early and mature malaria parasites, Biomed. Signal Process. Control 94 (2024) 106289, http://dx.doi.org/10.1016/j.bspc.2024.106289, URL https://linkinghub.elsevier.com/retrieve/pii/S1746809424003471.

[22] L. Zedda, A. Loddo, C. Di Ruberto, YOLO-PAM: Parasite-attention-based model for efficient malaria detection, J. Imaging 9 (12) (2023) 266, http://dx.doi.org/10.3390/jimaging9120266, Number: 12 Publisher: Multidisciplinary Digital Publishing Institute. URL https://www.mdpi.com/2313-433X/9/12/266.

[23] L. Zedda, A. Loddo, C. Di Ruberto, A deep learning based framework for malaria diagnosis on high variation data set, in: Image Analysis and Processing - ICIAP 2022 - 21st International Conference, Lecce, Italy, May 23-27, 2022, Proceedings, Part II, in: Lecture Notes in Computer Science, vol. 13232, Springer, 2022, pp. 358–370.

[24] L. Zedda, A. Loddo, C. Di Ruberto, A deep learning based framework for malaria diagnosis on high variation data set, in: Image Analysis and Processing - ICIAP 2022 - 21st International Conference, Lecce, Italy, May 23-27, 2022, Proceedings, Part II, in: Lecture Notes in Computer Science, vol. 13232, Springer, 2022, pp. 358–370.

[25] O.S. Zhao, N. Kolluri, A. Anand, N. Chu, R. Bhavaraju, A. Ojha, S. Tiku, D. Nguyen, R. Chen, A. Morales, et al., Convolutional neural networks to automate the screening of malaria in low-resource countries, PeerJ 8 (2020) e9674.

[26] S. Chibuta, A.C. Acar, Real-time malaria parasite screening in thick blood smears for low-resource setting, J. Digit. Imaging 33 (3) (2020) 763–775.

[27] V. Ljosa, K. Sokolnicki, A. Carpenter, Annotated high-throughput microscopy image sets for validation, Nature Methods 9 (7) (2012) 637, http://dx.doi.org/10.1038/nmeth.2083, arXiv:https://www.nature.com/articles/nmeth.2083.pdf. URL https://www.nature.com/articles/nmeth.2083.

[28] J.A. Quinn, A. Andama, I. Munabi, F.N. Kiwanuka, Automated blood smear analysis for mobile malaria diagnosis, in: Mobile Point-of-Care Monitors and Diagnostic Device Design, Vol. 31, CRC Press, 2014, p. 115.

[29] F. Abdurahman, K.A. Fante, M. Aliy, Malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models, BMC Bioinform. 22 (1) (2021) 112, http://dx.doi.org/10.1186/s12859-021-04036-4.

[30] A. Loddo, C.D. Ruberto, M. Kocher, G. Prod'Hom, MP-IDB: the malaria parasite image database for image processing and analysis, in: N. Leporé, J. Brieva, E. Romero, D. Racoceanu, L. Joskowicz (Eds.), Processing and Analysis of Biomedical Information - First International SIPAIM Workshop, SaMBa 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Revised Selected Papers, in: Lecture Notes in Computer Science, vol. 11379, Springer, 2018, pp. 57–65, http://dx.doi.org/10.1007/978-3-030-13835-6_7.

[31] D. Setyawan, R. Wardoyo, M. Wibowo, E. Murhandarwati, Classification of plasmodium falciparum based on textural and morphological features, Int. J. Electr. Comput. Eng. 12 (5) (2022) 5036–5048.

[32] K. delas Peñas, P.T. Rivera, P.C.N. Jr., Malaria parasite detection and species identification on thin blood smears using a convolutional neural network, in: P. Bonato, H. Wang (Eds.), Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2017, Philadelphia, PA, USA, July 17-19, 2017, IEEE Computer Society / ACM, 2017, pp. 1–6, http://dx.doi.org/10.1109/CHASE.2017.51.

[33] A. Rahman, H. Zunair, T.R. Reme, M.S. Rahman, M. Mahdy, A comparative analysis of deep learning architectures on high variation malaria parasite classification dataset, Tissue Cell 69 (2021) 101473.

[34] D.R. Loh, W.X. Yong, J. Yapeter, K. Subburaj, R. Chandramohanadas, A deep learning approach to the screening of malaria infection: Automated and rapid cell counting, object detection and instance segmentation using Mask R-CNN, Comput. Med. Imaging Graph. 88 (2021) 101845.

[35] C. Lin, H. Wu, Z. Wen, J. Qin, Automated malaria cells detection from blood smears under severe class imbalance via importance-aware balanced group softmax, in: M. de Bruijne, P.C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, C. Essert (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2021 - 24th International Conference, Strasbourg, France, September 27 - October 1, 2021, Proceedings, Part VIII, in: Lecture Notes in Computer Science, vol. 12908, Springer, 2021, pp. 455–465.

[36] A. Acherar, I. Tantaoui, M. Thellier, A. Lampros, R. Piarroux, X. Tannier, Real-life evaluation of deep learning models trained on two datasets for Plasmodium falciparum detection with thin blood smear images at 500x magnification, Inform. Med. Unlocked 35 (2022) 101132.

[37] W. Silka, M. Wieczorek, J. Silka, M. Wozniak, Malaria detection using advanced deep learning architecture, Sensors 23 (3) (2023) 1501, http://dx.doi.org/10.3390/S23031501.

[38] S.A. Kumar, M.K. Muchahari, S. Poonkuntran, L.S. Kumar, R.K. Dhanaraj, P. Karthikeyan, Application of hybrid capsule network model for malaria parasite detection on microscopic blood smear images, Multimedia Tools Appl. (2024) 1–27.

[39] P. Krishnadas, K. Chadaga, N. Sampathila, S. Rao, S.K. S., S. Prabhu, Classification of malaria using object detection models, Informatics 9 (4) (2022) http://dx.doi.org/10.3390/informatics9040076, URL https://www.mdpi.com/2227-9709/9/4/76.

[40] S. Mukherjee, S. Chatterjee, O. Bandyopadhyay, A. Biswas, Detection of malaria parasites in thin blood smears using CNN-based approach, in: J.K. Mandal, I. Mukherjee, S. Bakshi, S. Chatterji, P.K. Sa (Eds.), Computational Intelligence and Machine Learning, Springer Singapore, Singapore, 2021, pp. 19–27.

[41] A. Nautre, H.A. Nugroho, E.L. Frannita, R. Nurfauzi, Detection of malaria parasites in thin red blood smear using a segmentation approach with U-Net, in: 2020 3rd International Conference on Biomedical Engineering, IBIOMED, 2020, pp. 55–59, http://dx.doi.org/10.1109/IBIOMED50285.2020.9487603.

[42] A. Nanoti, S. Jain, C. Gupta, G. Vyas, Detection of malaria parasite species and life cycle stages using microscopic images of thin blood smear, in: 2016 International Conference on Inventive Computation Technologies, ICICT, Vol. 1, 2016, pp. 1–6, http://dx.doi.org/10.1109/INVENTIVE.2016.7823258.

[43] E. Var, F. Boray Tek, Malaria parasite detection with deep transfer learning, in: 2018 3rd International Conference on Computer Science and Engineering, UBMK, 2018, pp. 298–302, http://dx.doi.org/10.1109/UBMK.2018.8566549.

[44] S.S. Abbas, T.M. Dijkstra, Detection and stage classification of Plasmodium falciparum from images of Giemsa stained thin blood films using random forest classifiers, Diagn. Pathol. 15 (2020) 1–11.

[45] H.A.H. Chaudhry, M.S. Farid, A. Fiandrotti, M. Grangetto, A lightweight deep learning architecture for malaria parasite-type classification and life cycle stage detection, Neural Comput. Appl. (2024) 1–11.

[46] Q.A. Arshad, M. Ali, S. Hassan, C. Chen, A. Imran, G. Rasul, W. Sultani, A dataset and benchmark for malaria life-cycle classification in thin blood smear images, Neural Comput. Appl. 34 (6) (2022) 4473–4485, http://dx.doi.org/10.1007/s00521-021-06602-6.

[47] J. Hung, A. Goodman, S. Lopes, G. Rangel, D. Ravel, F.T.M. Costa, M. Duraisingh, M. Marti, A.E. Carpenter, Applying faster R-CNN for object detection on malaria images, 2018, CoRR abs/1804.09548. arXiv:1804.09548. URL http://arxiv.org/abs/1804.09548.

[48] R.R. Manku, A. Sharma, A. Panchbhai, Malaria detection and classificaiton, 2020, CoRR abs/2011.14329. arXiv:2011.14329. URL https://arxiv.org/abs/2011.14329.

[49] A. Loddo, C. Fadda, C.D. Ruberto, An empirical evaluation of convolutional networks for malaria diagnosis, J. Imaging 8 (3) (2022) 66, http://dx.doi.org/10.3390/jimaging8030066.

[50] N. Sengar, R. Burget, M. Dutta, A vision transformer based approach for analysis of plasmodium vivax life cycle for malaria prediction using thin blood smear microscopic images, Comput. Methods Programs Biomed. 224 (2022) 106996, http://dx.doi.org/10.1016/j.cmpb.2022.106996.

[51] S. Li, Z. Du, X. Meng, Y. Zhang, Multi-stage malaria parasite recognition by deep learning, GigaScience 10 (6) (2021) giab040, http://dx.doi.org/10.1093/gigascience/giab040, arXiv:https://academic.oup.com/gigascience/article-pdf/10/6/giab040/38673958/giab040.pdf.

[52] F. Yang, N. Quizon, H. Yu, K. Silamut, R.J. Maude, S. Jaeger, S. Antani, Cascading YOLO: automated malaria parasite detection for Plasmodium vivax in thin blood smears, in: H.K. Hahn, M.A. Mazurowski (Eds.), Medical Imaging 2020: Computer-Aided Diagnosis, in: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 11314, 2020, p. 113141Q, http://dx.doi.org/10.1117/12.2549701.

[53] M.B. Jamshidi, A. Lalbakhsh, J. Talla, Z. Peroutka, S. Roshani, V. Matousek, S. Roshani, M. Mirmozafari, Z. Malek, L. La Spada, A. Sabet, M. Dehghani, M. Jamshidi, M.M. Honari, F. Hadjilooei, A. Jamshidi, P. Lalbakhsh, H. Hashemi-Dezaki, S. Ahmadi, S. Lotfi, Deep learning techniques and COVID-19 drug discovery: Fundamentals, state-of-the-art and future directions, in: I. Arpaci, M. Al-Emran, M. A. Al-Sharafi, G. Marques (Eds.), Emerging Technologies During the Era of COVID-19 Pandemic, Springer International Publishing, Cham, 2021, pp. 9–31, http://dx.doi.org/10.1007/978-3-030-67716-9_2.

[54] M. Jamshidi, A. Lalbakhsh, J. Talla, Z. Peroutka, F. Hadjilooei, P. Lalbakhsh, M. Jamshidi, L.L. Spada, M. Mirmozafari, M. Dehghani, A. Sabet, S. Roshani, S. Roshani, N. Bayat-Makou, B. Mohamadzade, Z. Malek, A. Jamshidi, S. Kiani, H. Hashemi-Dezaki, W. Mohyuddin, Artificial intelligence and COVID-19: Deep learning approaches for diagnosis and treatment, IEEE Access 8 (2020) 109581–109595, http://dx.doi.org/10.1109/ACCESS.2020.3001973, URL https://ieeexplore.ieee.org/document/9115663/.

[55] L. Lamm, R.D. Righetto, W. Wietrzynski, M. Pöge, A. Martinez-Sanchez, T. Peng, B.D. Engel, MemBrain: A deep learning-aided pipeline for detection of membrane proteins in Cryo-electron tomograms, Comput. Methods Programs Biomed. 224 (2022) 106990, http://dx.doi.org/10.1016/j.cmpb.2022.106990, URL https://www.sciencedirect.com/science/article/pii/S0169260722003728.

[56] L. Lamm, S. Zufferey, R.D. Righetto, W. Wietrzynski, K.A. Yamauchi, A. Burt, Y. Liu, H. Zhang, A. Martinez-Sanchez, S. Ziegler, F. Isensee, J.A. Schnabel, B.D. Engel, T. Peng, MemBrain v2: an end-to-end tool for the analysis of membranes in cryo-electron tomography, BioRxiv (2024) http://dx.doi.org/10.1101/2024.01.05.574336, arXiv:https://www.biorxiv.org/content/early/2024/01/05/2024.01.05.574336.full.pdf. URL https://www.biorxiv.org/content/early/2024/01/05/2024.01.05.574336.

[57] L. Sun, Z. Fan, Y. Huang, J. Paisley, Compressed sensing MRI using a recursive dilated network, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, 2018, http://dx.doi.org/10.1609/aaai.v32i1.11869.

[58] O. Attallah, Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning, Comput. Biol. Med. 178 (2024) 108798.

[59] K. Mao, X. Jing, G. Wang, Y. Chang, J. Liu, Y. Zhao, S. Yu, J. Liu, A novel open-source CADs platform for 3D CT pulmonary analysis, Comput. Biol. Med. 169 (2024) 107878.

[60] Z. Yu, X. Li, J. Li, W. Chen, Z. Tang, D. Geng, HSA-net with a novel CAD pipeline boosts both clinical brain tumor MR image classification and segmentation, Comput. Biol. Med. 170 (2024) 108039.

[61] Q. Hu, C. Chen, S. Kang, Z. Sun, Y. Wang, M. Xiang, H. Guan, L. Xia, S. Wang, Application of computer-aided detection (CAD) software to automatically detect nodules under SDCT and LDCT scans with different parameters, Comput. Biol. Med. 146 (2022) 105538.

[62] M. Chossegros, X. Tannier, D. Stockholm, Improving interpretability of leucocyte classification with multimodal network, Stud. Health Technol. Inform. 316 (2024) 1098–1102, http://dx.doi.org/10.3233/SHTI240602.

[63] M. Chossegros, F. Delhommeau, D. Stockholm, X. Tannier, Improving the generalizability of white blood cell classification with few-shot domain adaptation, J. Pathol. Inform. (2024) 100405.

[64] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proc. of the 25th International Conference on Neural Information Processing Systems, NIPS '12, Vol. 1, 2012, pp. 1097–1105.

[65] Q.A. Arshad, M. Ali, S. Hassan, C. Chen, A. Imran, G. Rasul, W. Sultani, A dataset and benchmark for malaria life-cycle classification in thin blood smear images, Neural Comput. Appl. 34 (6) (2022) 4473–4485.

[66] S. Marletta, V. L'Imperio, A. Eccher, P. Antonini, N. Santonicco, I. Girolami, A.P. Dei Tos, M. Sbaraglia, F. Pagni, M. Brunelli, et al., Artificial intelligence-based tools applied to pathological diagnosis of microbiological diseases, Pathol.-Res. Pr. (2023) 154362.

[67] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-Unet: Unet-like pure transformer for medical image segmentation, 2023, http://dx.doi.org/10.1007/978-3-031-25066-8_9, URL https://link.springer.com/chapter/10.1007/978-3-031-25066-8_9.

[68] A. Loddo, L. Putzu, On the reliability of CNNs in clinical practice: a computer-aided diagnosis system case study, Appl. Sci. 12 (7) (2022) 3269.

[69] P. Manescu, C. Bendkowski, R. Claveau, M. Elmi, B.J. Brown, V. Pawar, M.J. Shaw, D. Fernandez-Reyes, A weakly supervised deep learning approach for detecting malaria and sickle cells in blood films, in: A.L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M.A. Zuluaga, S.K. Zhou, D. Racoceanu, L. Joskowicz (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2020 - 23rd International Conference, Lima, Peru, October 4-8, 2020, Proceedings, Part V, in: Lecture Notes in Computer Science, vol. 12265, Springer, 2020, pp. 226–235.

[70] A. Koirala, M. Jha, S. Bodapati, A. Mishra, G. Chetty, P.K. Sahu, S. Mohanty, T.K. Padhan, J. Mattoo, A. Hukkoo, Deep learning for real-time malaria parasite detection and counting using YOLO-mp, IEEE Access 10 (2022) 102157–102172.

[71] T. Fatima, M.S. Farid, Automatic detection of Plasmodium parasites from microscopic blood images, J. Parasit. Dis. 44 (1) (2020) 69–78.

[72] M. Fu, K. Wu, Y. Li, L. Luo, W. Huang, Q. Zhang, An intelligent detection method for plasmodium based on self-supervised learning and attention mechanism, Front. Med. (Lausanne) 10 (2023) 1117192, http://dx.doi.org/10.3389/fmed.2023.1117192.

[73] A.M. Rekavandi, S. Rashidi, F. Boussaid, S. Hoefs, E. Akbas, M. bennamoun, Transformers in small object detection: A benchmark and survey of state-of-the-art, 2023, arXiv:2309.04902. URL http://arxiv.org/abs/2309.04902.

[74] X. Zhu, S. Lyu, X. Wang, Q. Zhao, TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2021, pp. 2778–2788.

[75] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. von Luxburg, S. Bengio, H.M. Wallach, R. Fergus, S.V.N. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, 2017, pp. 5998–6008.

[76] S. Woo, J. Park, J. Lee, I.S. Kweon, CBAM: convolutional block attention module, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII, in: Lecture Notes in Computer Science, vol. 11211, Springer, 2018, pp. 3–19, http://dx.doi.org/10.1007/978-3-030-01234-2_1.

[77] Y. Liu, Z. Shao, Y. Teng, N. Hoffmann, NAM: normalization-based attention module, 2021, CoRR abs/2111.12419. arXiv:2111.12419. URL https://arxiv.org/abs/2111.12419.

[78] A. Diker, An efficient model of residual based convolutional neural network with Bayesian optimization for the classification of malarial cell images, Comput. Biol. Med. (2022) 105635.

[79] N. Sengar, R. Burget, M.K. Dutta, A vision transformer based approach for analysis of plasmodium vivax life cycle for malaria prediction using thin blood smear microscopic images, Comput. Methods Programs Biomed. 224 (2022) 106996, http://dx.doi.org/10.1016/j.cmpb.2022.106996.

[80] H. Guan, M. Liu, Domain adaptation for medical image analysis: A survey, IEEE Trans. Biomed. Eng. 69 (3) (2022) 1173–1185, http://dx.doi.org/10.1109/TBME.2021.3117407.

[81] W.M. Kouw, M. Loog, A review of domain adaptation without target labels, 2019, arXiv:1901.05335. URL http://arxiv.org/abs/1901.05335.

[82] I.R. Dave, T.d. Blegiers, C. Chen, M. Shah, Codamal: Contrastive domain adaptation for malaria detection in low-cost microscopes, in: 2024 IEEE International Conference on Image Processing, ICIP, 2024, pp. 3848–3853.

[83] C.W. Pirnstill, G.L. Coté, Malaria diagnosis using a mobile phone polarized microscope, Sci. Rep. 5 (1) (2015) 13368, http://dx.doi.org/10.1038/srep13368.

[84] F. Yang, M. Poostchi, H. Yu, Z. Zhou, K. Silamut, J. Yu, R.J. Maude, S. Jäger, S.K. Antani, Deep learning for smartphone-based malaria parasite detection in thick blood smears, IEEE J. Biomed. Heal. Inform. 24 (5) (2020) 1427–1438.

[85] W. Sultani, W. Nawaz, S. Javed, M.S. Danish, A. Saadia, M. Ali, Towards low-cost and efficient malaria detection, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, IEEE, 2022, pp. 20655–20664, http://dx.doi.org/10.1109/CVPR52688.2022.02003.

[86] W. Wang, J. Dai, Z. Chen, Z. Huang, Z. Li, X. Zhu, X. Hu, T. Lu, L. Lu, H. Li, X. Wang, Y. Qiao, InternImage: Exploring large-scale vision foundation models with deformable convolutions, 2022, http://dx.doi.org/10.48550/arXiv.2211.05778, CoRR abs/2211.05778. arXiv:2211.05778.

[87] Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A ConvNet for the 2020s, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, IEEE, 2022, pp. 11966–11976, http://dx.doi.org/10.1109/CVPR52688.2022.01167.

[88] M. Caron, H. Touvron, I. Misra, H. Jegou, J. Mairal, P. Bojanowski, A. Joulin, Emerging properties in self-supervised vision transformers, 2021, arXiv:2104.14294.

[89] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, 2020, CoRR abs/2010.11929. arXiv:2010.11929. URL https://arxiv.org/abs/2010.11929.

[90] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: 2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021, IEEE, 2021, pp. 9992–10002.

[91] S. Ren, K. He, R.B. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada, 2015, pp. 91–99.

[92] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S.E. Reed, C. Fu, A.C. Berg, SSD: single shot multibox detector, in: Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, the Netherlands, October 11-14, 2016, Proceedings, Part I, in: Lecture Notes in Computer Science, vol. 9905, Springer, 2016, pp. 21–37.

[93] T. Lin, P. Dollár, R.B. Girshick, K. He, B. Hariharan, S.J. Belongie, Feature pyramid networks for object detection, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, IEEE Computer Society, 2017, pp. 936–944.

[94] J. Redmon, S.K. Divvala, R.B. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, IEEE Computer Society, 2016, pp. 779–788.

[95] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, IEEE Computer Society, 2017, pp. 6517–6525.

[96] J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, 2018, CoRR abs/1804.02767.

[97] A. Bochkovskiy, C. Wang, H.M. Liao, YOLOv4: Optimal speed and accuracy of object detection, 2020, CoRR abs/2004.10934.

[98] H. Zhou, F. Jiang, H. Lu, SSDA-YOLO: semi-supervised domain adaptive YOLO for cross-domain object detection, 2022, CoRR abs/2211.02213.

[99] Z. Zou, K. Chen, Z. Shi, Y. Guo, J. Ye, Object detection in 20 years: A survey, Proc. IEEE (2023) 1–20, http://dx.doi.org/10.1109/JPROC.2023.3238524.

[100] T. Lin, M. Maire, S.J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft COCO: common objects in context, in: Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V, in: Lecture Notes in Computer Science, vol. 8693, Springer, 2014, pp. 740–755.

[101] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916.

[102] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation, in: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018, Computer Vision Foundation / IEEE Computer Society, 2018, pp. 8759–8768.

[103] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.

[104] M. Guo, T. Xu, J. Liu, Z. Liu, P. Jiang, T. Mu, S. Zhang, R.R. Martin, M. Cheng, S. Hu, Attention mechanisms in computer vision: A survey, Comput. Vis. Media 8 (3) (2022) 331–368.

[105] S. Woo, J. Park, J. Lee, I.S. Kweon, CBAM: convolutional block attention module, in: Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII, in: Lecture Notes in Computer Science, vol. 11211, Springer, 2018, pp. 3–19.

[106] y. liu, Z. Shao, y. Teng, N. Hoffmann, NAM: Normalization-based attention module, in: NeurIPS 2021 Workshop on ImageNet: Past, Present, and Future, 2021, p. 8, URL https://openreview.net/forum?id=AaTK_ESdkjg.

[107] R. Padilla, S.L. Netto, E.A.B. da Silva, A survey on performance metrics for object-detection algorithms, in: 2020 International Conference on Systems, Signals and Image Processing, IWSSIP 2020, Niterói, Brazil, July 1-3, 2020, IEEE, 2020, pp. 237–242.

[108] G. Jocher, A. Chaurasia, J. Qiu, Ultralytics YOLOv8, 2023, URL https://github.com/ultralytics/ultralytics.

[109] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft COCO: Common objects in context, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014, pp. 740–755.

[110] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, B. Guo, Swin transformer V2: scaling up capacity and resolution, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, IEEE, 2022, pp. 11999–12009, http://dx.doi.org/10.1109/CVPR52688.2022.01170.

[111] S. Chen, S. Zhao, C. Huang, An automatic malaria disease diagnosis framework integrating blockchain-enabled cloud-edge computing and deep learning, IEEE Internet Things J. 10 (24) (2023) 21544–21553, http://dx.doi.org/10.1109/JIOT.2023.3304526.

[112] T. Mutabazi, E. Arinaitwe, A. Ndyabakira, E. Sendaula, A. Kakeeto, P. Okimat, P. Orishaba, S.P. Katongole, A. Mpimbaza, P. Byakika-Kibwika, C. Karamagi, J.N. Kalyango, M.R. Kamya, G. Dorsey, J.I. Nankabirwa, Assessment of the accuracy of malaria microscopy in private health facilities in Entebbe Municipality, Uganda: a cross-sectional study, Malar. J. 20 (1) (2021) 250, http://dx.doi.org/10.1186/s12936-021-03787-y.

[113] T. Dao, D.Y. Fu, S. Ermon, A. Rudra, C. Ré, FlashAttention: Fast and memory-efficient exact attention with IO-awareness, in: Advances in Neural Information Processing Systems, NeurIPS, 2022.

[114] T. Dao, FlashAttention-2: Faster attention with better parallelism and work partitioning, in: International Conference on Learning Representations, ICLR, 2024.