Full Length Article

Uncertainty-guided attention learning for malaria parasite detection in thick blood smears

Hao Xiong [a,*], Zhiyong Wang [b], Roneel V. Sharan [a,c], Shlomo Berkovsky [a]

[a] *Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, 2109, Sydney, NSW, Australia*
[b] *School of Computer Science, The University of Sydney, 2006, Sydney, NSW, Australia*
[c] *School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ, Colchester, United Kingdom*

A B S T R A C T

Malaria may seriously threaten an individual's health and wellbeing, and early screening is pivotal for timely treatment and recovery. In malaria screening, thick blood smears are exploited to count the parasites and assess the severity of the disease. Parasites are tiny objects that can be found in high resolution blood smear images, which renders them difficult for detection. Other than using object detection based methods, prior works also applied image classification techniques to this problem. They first extracted image patches from blood smears as parasite candidates and then utilized convolutional neural networks to classify these patches as parasites or non-parasites. However, these approaches overlook the fact that the blood smear images may contain noises, errors, and background artifacts, which introduces uncertainty and makes the model predictions less stable. In this work, we propose an uncertainty-guided attention learning based network for malaria parasite detection from thick blood smears, which incorporates pixel attention mechanism to identify more fine-grained and pixel-wise informative features, to improve the classification capability of our model. We further put uncertainty estimation on channels of the feature map to guide pixel attention learning, such that the features from channels with higher uncertainty are considered unreliable and are thus restrictively exploited by pixel attention learning. To estimate channel-wise uncertainty, we introduce the Bayesian channel attention, which reformulates the traditional channel attention under the Bayesian framework. As a result, it denotes channel uncertainties with estimated variances that guide the pixel attention learning. We compared to several state-of-the-art baselines on two public datasets using parasite-level and patient-level evaluations. The proposed method demonstrates superior performance with respect to most metrics on two datasets, especially achieving highest average precision (AP) scores in both parasite and patient-level scenarios.

1. Introduction

Malaria is a life threatening disease caused by parasites that are transmitted to people by the bite of infected mosquitoes. The 2022 malaria report of the World Health Organization suggests that 247 million malaria cases were detected worldwide and 619,000 deaths were reported in 2022 (Organization, 2022). Although malaria is curable, improper diagnosis and treatment may lead to various complications and even death. Owing to low cost and accessibility, microscopic examination of thick and thin blood smears is currently the gold standard for malaria diagnosis (Makhija et al., 2015). Blood smears are utilized to detect the presence of malaria parasites in blood, and the parasite density in blood smears is a key index for diagnosing malaria and quantifying its severity. Experts normally examine the thick and thin smears for parasite detection; yet visual examination of blood smears is a tedious procedure. Also, well-trained experts are scarce, especially in malaria-endemic and resource-constrained areas, while intensive loads may cause human fatigue and result in inaccurate diagnosis leading to poor clinical decision-making. Thus, an automatic and reliable malaria diagnosis tool is urgently needed.

In this work, we exploit thick blood smear images for malaria screening. Compared to thin blood smears, thick smears provide sufficient volumes of blood and facilitate more reliable diagnosis of low parasite density cases, with approximately 11 times higher sensitivity (Warhurst & Williams, 1996). A thick blood smear is produced by spreading a drop of blood on a slide which is then dried and stained with Romanovsky-type stain (Organization, 2020). Though thick blood smears are more sensitive/informative than thin smears, the parasites in thick smears

---

* Corresponding author.

E-mail addresses: hao.xiong@mq.edu.au (H. Xiong), zhiyong.wang@sydney.edu.au (Z. Wang), roneel.sharan@essex.ac.uk (R.V. Sharan), shlomo.berkovsky@mq.edu.au (S. Berkovsky).

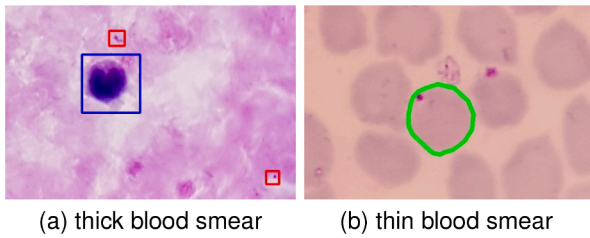(a) thick blood smear   (b) thin blood smear

Fig. 1. Example of thick and thin blood smears. (a) The red bounding box contains the parasite that appears as a small purple disk. The blue box refers to the white blood cell (WBC), which has similar color to parasite but looks much larger. (b) the green circle contains the parasitized red blood cell which appears more visible and easier to detect than parasites in thick blood smear.

are stained as small, round, purple disks that cannot always be effectively identified. A visual comparison between thick and thin blood smears is shown in Fig. 1. As can be seen, the red blood cells in thin blood smear appear to be much larger and are thus easier to detect than those stained small parasites in thick blood smears. Besides that, some non-parasite components of the thick film also absorb the stain and appear similar to parasites. These components, or distractors, induce background noises and affect the identification of parasites. Thus, we focus our work on identification of malaria parasites in thick smear images.

Traditional parasite detection methods tend to extract low-level features, which fed into classifiers for automatic detection (Elter et al., 2011a; Purnama et al., 2013; Quinn et al., 2014; Rosado et al., 2016b; Yunda, 2011). Such hand-crafted or low-level features can hardly ensure sufficiently high classification accuracy. Due to recent advancement of deep learning, a large number of deep learning based methods have been proposed to detect objects in images (Dai et al., 2016; Girshick, 2015; Girshick et al., 2014; Redmon et al., 2016; Ren et al., 2017). Based on these, Lin et al. proposed a malaria cell detection method to solve class imbalance problem between healthy and infected red blood cells (Delahunt et al., 2015). In fact, parasites are small objects found in high resolution thick blood smear images; therefore, the above-mentioned object detection methods are not sufficiently effective to identify tiny malaria parasites. In addition to object detection, alternative methods first extracted a number of patches from blood smear images as parasite candidates and then classified these patches to identify parasites (Manescu et al., 2020a; Mehanian et al., 2017a; Quinn et al., 2016; Yang et al., 2020). For example, Petru et al. proposed a weakly supervised approach for malaria parasite detection (Manescu et al., 2020a). Instead of labeling all the parasites, they only annotated each blood film as malaria positive or negative. While this facilitated annotation, the labels were weak annotations, rendering inferior performance compared to fully supervised method. Besides that, Yang et al. proposed a fully supervised method for malaria parasite detection Yang et al., 2020. They preselected parasite candidates, which included many possible parasite instances, and fed them into a convolutional neural network (CNN) to classify true parasites and achieved promising results. They also demonstrated that this model could be deployed on a mobile phone for effective malaria parasites detection.

However, the above approaches largely ignore the fact that the acquired images are potentially noisy, may introduce uncertainty into the learned features, and make the model predictions less reliable. In this work, we propose an uncertainty-guided attention learning based network to address the feature uncertainty issue caused by the data noises. To achieve this, we combine attention learning with a CNN-based network under a unified Bayesian framework that introduces a Gaussian distribution to the attention learning. The variance of this Gaussian distribution reflects channel-wise uncertainty (reliability) of features, and is defined as a function of input data and network weights. By doing that, we ensure that the uncertainty estimation is input-dependent, to miti-

gate the data noises that may vary across the input instances. As a result, the channel features with higher variance are considered less reliable and are restrictively utilized by our model. Upon obtaining the variance, we normalize it within the [0,1] range to represent the uncertainty in a soft manner, such that channel features with higher uncertainty have lower weights and contribute less, and vice versa. Based on the channel features with estimated uncertainty, we apply pixel attention to further identify important pixel-wise information from the reliable channel features by also treating different pixels in channel features in different ways. As a result, the important pixel-wise information extracted from reliable channel features effectively facilitate our task of parasite identification. Experimental results on two publicly available *CMM* and *Thick Smear 150* malaria parasite detection datasets demonstrate the superiority of the proposed approach over state-of-the-art baseline methods.

It is worth to note that many recent attention-based approaches emphasize capturing attended features or extracting long range information. Rather, our attention learning aims to identify input-level reliable features. Here, our contributions can be summarized as follows:

- We propose a novel Bayesian framework that ensures input-level uncertainty estimation to mitigate variable noises in the input data. This is achieved by modelling the variance estimation as an input-dependent function.
- To the best of our knowledge, our work is the first to estimate feature uncertainties, exploiting the more reliable features with estimated uncertainties for the malaria parasite detection task.
- Our work offers a promising alternative to the manual parasite detection, which is particularly important in malaria-endemic areas, often suffering from a shortage of experienced parasitologists. It also allows to estimate the malaria severity and prioritize patient treatment, which can help in under-developed malaria-endemic areas.

2. Related work

In this section, we provide an overview of existing methods relevant to our study. We first discuss traditional and deep learning based methods used for malaria parasites detection, and then examine uncertainty estimation in medical imaging, focusing on techniques designed to quantify model confidence and enhance the reliability of deep learning models in clinical applications.

2.1. Traditional methods

Traditional parasite detection methods exploit segmentation techniques to identify parasites. To achieve that, these works primarily rely on setting thresholds and morphological operations. For instance, since malaria parasites tend to have darker intensities, Hanif et al. (2011) empirically set a threshold to filter out darker regions as parasites. To further enhance the saliency of parasites and facilitate their identification, they applied the intensity-stretching method which is a linear mapping function commonly used to enhance the brightness and contrast level of the image. It consists of two processes - dark stretching process and bright compression process, for which the former stretches the range of image values below a pre-set threshold, while the later compresses the image values greater than that threshold. By doing so, the parasite regions in the image become more distinguishable. Likewise, Dave and Upla (2017) proposed a histogram-based adaptive thresholding to detect malaria parasites in thin and thick blood smears. Chakrabortya et al. (2015) incorporated color information into a morphological segmentation method to identify malaria parasites. Unlike the above works, Kaewkamnerd et al. (2011) proposed to identify malaria parasites in HSV images. This is primarily because the V channel in HSV image efficiently represent the lightness (brightness) of image compared to the RGB image, and can better separate parasites from background. Hence,

they first converted RGB blood smear images into HSV images, and then calculated the histogram of the V channel in the HSV image. Finally, they used an adaptive threshold on the V-channel histogram to extract parasites that normally have higher values in histogram. In addition to these segmentation methods, there are other hand-crafted features based approaches (Abdul Nasir et al., 2012; Elter et al., 2011b; Park et al., 2016; Rosado et al., 2016a), for which hand-crafted features are extracted from parasite candidates and then fed into classifiers, such as support vector machine (SVM), linear regression (LR), K-means clustering algorithm, to identify malaria parasites. However, it turns out that both threshold based segmentation and hand-crafted features are low-level image and morphological techniques that struggle to ensure parasite detection accuracy.

2.2. Deep learning based methods

In recent years, deep learning has been successful in many scenarios and applications. In malaria parasites detection, some works (Gopakumar et al., 2018; Vijayalakshmi & Rajesh, 2020) utilized CNN models to extract features which were shown to be more representative and informative than the hand-crafted ones. Afterwards, these CNN extracted features were fed into classical machine learning methods, such as SVM, to allow an accurate parasite classification. By contrast, other methods (Abdurahman et al., 2021; Doering et al., 2020; Lin et al., 2021; Loddo et al., 2018; Mehanian et al., 2017b) developed end-to-end networks to detect malaria parasites in thick blood smears. They considered images as input and produced the identified parasites directly. These works exploited common state-of-the-arts methods; for example, Abdurahman et al. (2021), Lin et al. (2021) were object detection based methods. Due to the class imbalance between parasite infected cell and other cells, the learned model may be biased to identify the cells that have more training samples. Therefore, Lin et al. (2021) proposed to alleviate the class imbalance issue by exploiting a relation module to learn the relationships between the different cells. The work of Abdurahman et al. (2021) was a Yolov4 based network that modified the traditional object detection Yolov4 by extending feature scales and adding more detection layers to extract better features. Other works Guemas et al., 2024; Sukumarran et al., 2024a,b; Zhao et al., 2020 utilized thin blood smear images to detect malaria-infected cells. For malaria-infected cells detection, they exploited the commonly-used object detection methods such as RT-DETR and YOLO. Notably, Sukumarran et al. (2024a) modified conventional YOLOv4 using direct layer pruning and backbone replacement, and in Sukumarran et al. (2024b) they further classified the detected infected cells into four Plasmodium species. Other than thick and thin blood smears, Dev et al. (2024), Madhu et al. (2023), Mujahid et al. (2024) also explored deep learning based classification approach to identify malaria parasites in the red blood cell image, for which the image had only one red blood cell and the task was to classify it as infected or not. For instance, in Dev et al. (2024), Madhu et al. (2023) the authors applied existing CNN-RNN based and inception based networks, while Mujahid et al. (2024) proposed a novel EfficientNet for parasite cell classification.

The success of deep learning methods largely relies on their representation of the extracted features. However, none of the above methods considers feature uncertainty introduced by potential data noises that may cause performance degradation in parasites detection.

2.3. Uncertainty estimation in medical imaging

Existing uncertainty studies of medical imaging mainly focus on quantifying the uncertainty of predictions, such that the predicted output is associated with uncertainty estimation to provide a confidence evaluation of the prediction. As a result, the model with uncertainty estimation becomes more reliable and trustworthy to allow better clinical decisions.

A simple approach to generate uncertainty is exploiting the deterministic network, such as the evidential deep learning methods (Amini et al., 2020; Sensoy et al., 2018), to output both the predictions and uncertainty scores. Based on deterministic uncertainty estimation, Amersfoort et al. (2020) proposed a radial basis function based network for medical image segmentation. In Huang et al. (2022), Huang *et al.* developed a belief function at each modality voxel to model uncertainty, and then harnessed Dempster's rule for multi-modal medical image segmentation. Meanwhile, ensemble methods (Ashukha et al., 2019; Lakshminarayanan et al., 2017; Shen & Cremers, 2022) are alternative approaches, which aggregate the predictions of multiple models with the variance of each prediction to estimate uncertainty. For instance, Cao et al. (2020) designed an adaptive ensembling momentum map based on ensemble learning and proposed an uncertainty aware unsupervised loss for breast mass segmentation. Likewise, Mehrtash et al. (2020) proposed the multi-FCNs ensembling network to estimate uncertainty for medical image segmentation.

In addition to deterministic and ensemble methods, most existing medical imaging tasks rely on Bayesian neural networks to estimate uncertainty. For example, Leibig et al. (2017) detected diabetic retinopathy from fundus image using dropout based Bayesian uncertainty to provide confidence of the diagnosis. For skin lesion analysis, Molle et al. (2019) quantified the uncertainty of skin lesion classification using a novel Bayesian measure based on overlap of output distributions. Another work (Abdar et al., 2021) classifying skin cancer images introduced three uncertainty quantification methods, including Monte Carlo (MC) Dropout, ensemble MC and deep ensemble, into a novel hybrid dynamic Bayesian deep learning model based on a three branch decision theory.

In the field of Magnetic Resonance Image (MRI), Narnhofer et al. (2022) exploited the total deep variation regularizer for single and multi-coil undersampled MRI reconstruction, and quantified the uncertainty with a proposed Bayesian framework. Besides, Herzog et al. (2020) detected stroke lesion on 2D MR images by proposing a Bayesian convolutional neural network that can also generated an uncertainty estimation to facilitate the reliability of prediction. Similarly, Prince et al. (2023) quantified uncertainty for classification of Adamantinomatous Craniopharyngioma from preoperative MRI by exploiting the Variational Inference with elliptical slice sampling.

Other works also applied uncertainty estimations in the medical image segmentation, since in medical images some particular regions of the tissue may not be clear. For example, Kohl et al. (2018) proposed a probabilistic U-Net exploiting conditional variational autoencoder to learn a distribution over the output segmentation map. Consequently, their method was able to generate an unlimited number of hypotheses of segmentation maps on the lung abnormalities segmentation task. Based on probabilistic U-Net, subsequent works (Baumgartner et al., 2019; Gantenbein et al., 2020; Hu et al., 2019) further improved model efficiency and introduced epistemic uncertainty, and applied their methods in segmentation tasks on general medical images. Rather than generating multiple segmentation maps, Sedai et al. (2018) estimated a pixelwise uncertainty map along with the retinal layer segmentation. Here, the uncertainty map was utilized to identify erroneously segmented regions and thus facilitated downstream tasks. Besides, Tang et al. (2022) proposed an UG-Net that first generated a coarse segmentation map and a pixel-wise uncertainty map of the segmentation. The uncertainty map was utilized to refine the coarse segmentation map in an end-to-end manner.

The aforementioned methods tend to estimate uncertainty and quantify the reliability of predictions, such that subsequent clinical decisions can benefit from it. Unlike our method, none of these approaches, particularly in malaria parasite detection, focuses on learning the uncertainty of features and then utilizes the estimated uncertainty to mitigate the effects of unreliable features. To our knowledge, we are the first to estimate the uncertainty of the learned features, for which we harness the reliable features for subsequent detection of malaria parasites.

3. Method

3.1. Overview

The overall framework of our method is shown in Fig. 2. By using the iterative global minimum screening (IGMS) (Yang et al., 2020), we first select a large number of parasite candidates which are small image patches containing both parasites and distractors. We then recognize the true parasites out of these selected image patches using our proposed parasite classification network that includes two key complementary attention modules: (1) Bayesian Channel Attention that estimates uncertainty of feature channels, and (2) Uncertainty-Guided Pixel Attention that extracts important pixel-wise information from reliable channel features identified by (1). In the subsequent sections, we use $\mathbf{X} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}] \in \mathbb{R}^{N \times P}$ to denote N input image patches, each having P pixels. Their labels $\mathbf{Y} = [\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(N)}] \in \mathbb{R}^{N \times 1}$ indicate whether the image patch contains a parasite or not.

3.2. Parasites candidates preselection

The parasite candidates preselection aims to select a subset of most likely parasite candidates such that reducing the initial search space for parasite identification.

The candidates are selected based on the observation that the parasite pixels normally have lower intensities in a blood smear image. However, the WBCs appear to have darker intensities in addition to the parasites (as shown in Fig. 1). Therefore, the first step is to eliminate the distraction of WBCs before extracting parasite candidates. Here, we follow the methodology proposed by Yang et al. (2020), for which it consists of two main steps: WBC detection to remove WBCs and parasite candidate generation to localize regions of interest using lowest intensities. First, thick blood smear images are converted into grayscale images and then the Otsu's method (Otsu, 1979) is applied to obtain the binary mask of WBC, and subsequently WBC areas become easy to identify and remove.

After removing WBC regions, we apply parasite candidate generation using the iterative global minimum screening (IGMS) algorithm as described in Yang et al. (2020). Briefly, the parasites have low intensities; thus, the pixels having minimum intensity values are identified as parasite candidates. Once a pixel is identified as parasite candidate, a region centred at this pixel location with radius 22 is cropped. It is

noteworthy that the cropped image becomes circular with area outside circular region set to be black (shown as Selected Parasite Candidates in Fig. 2). In line with Yang et al. (2020), 500 image patches of size $44 \times 44 \times 3$ are iteratively selected as parasite candidates. More details of the above procedures can be found in Yang et al. (2020).

3.3. Proposed parasite classification network

We propose a parasite classification network to classify the extracted parasite candidates either as parasites or distractors. We exploit the CNN network originally proposed by the malaria parasite detection method (Yang et al., 2020) as our backbone. The network of Yang et al. (2020) is derived from VGG19 (Simonyan & Zisserman, 2015) and consists of two key parts: feature extraction and classification. The feature extraction part generates features that are utilized by the subsequent classification part for parasite identification. In this work, we aim to identify and exploit the reliable feature channels from the feature extraction part for more accurate parasite classification.

As illustrated in Fig. 3, our approach can be broken down into two stages: the first is *uncertainty estimation* that identifies reliable channel features, while the second exploits pixel attention to *extract important pixel-wise information* from the identified reliable channel features, in order to achieve more accurate and reliable parasite identification.

The output of feature extraction (from the 7th convolutional layer in Fig. 3) is of size $5 \times 5 \times 64$, where 64 refers to the number of channels. The output is considered as the final set of features fed into the subsequent fully-connected layers for classification. At the first stage, we perform channel-wise uncertainty estimation on each of these 64 channels. Consequently, the channel features with higher uncertainty are expected to contribute less to the final classification and vice versa. At the second stage, we substitute the Bayesian channel attention with uncertainty-guided pixel attention. Likewise, we insert pixel attention after the same feature map with 64 channels from last convolutional layer. With estimated channel-wise uncertainty of channel features, we limit the information from unreliable channels propagate into pixel attention learning. As a result, the informative pixel-wise features identified by our pixel attention are based on reliable channel features and is able to further improve the classification of parasites. It is also worth noting that pixel attention is able to estimate pixel-wise importance of the feature map and thus being more flexible and effective than channel attention.

3.4. Bayesian channel attention

In a deep learning network, each channel in the feature map cannot be treated equally. Unlike conventional channel attention mechanisms such as Dai et al. (2021), which assign each channel a single scalar weight to indicate its importance, our design explicitly captures the confidence level of these importance estimates, offering robustness in the presence of noisy or redundant features.

Channel attention estimates an attention weight for each feature channel indicating how important this channel is. However, it does not provide further information on the reliability of these features with the estimated attention weights. We assume that the collected data are inherently noisy. Hence, the feature map generated by the neural network from such data are noisy as well, and thus all the feature channels cannot be treated equally reliable. In addition, different input data have different noises, and are supposed to generate features with varied levels of reliability. Therefore, feature reliability should be dependent on the level of input noises. Here, we consider input-adaptive uncertainties when estimating the weights of channel features. Specifically, the channel features with less uncertainty are regarded as more reliable to be utilized for subsequent tasks and vice versa. We reformulate traditional channel attention as Bayesian channel attention to estimate channel-wise uncertainty. In next sections, we first introduce traditional
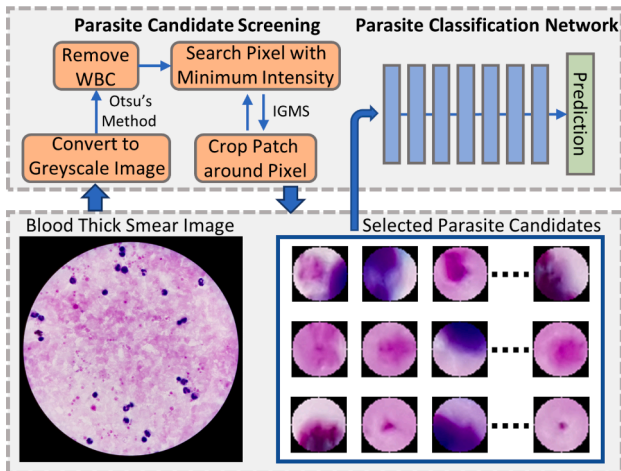

Fig. 2. The flowchart of our parasite identification pipeline. First, a blood thick smear image is processed by **Parasite Candidates Screening** to select a large number of image patches as parasite candidates. Then, our proposed **Parasite Classification Network** will identify true parasites from these candidates. The above blood thick smear images are from the dataset *Thick Smears 150* collected by Yang et al. (2020).
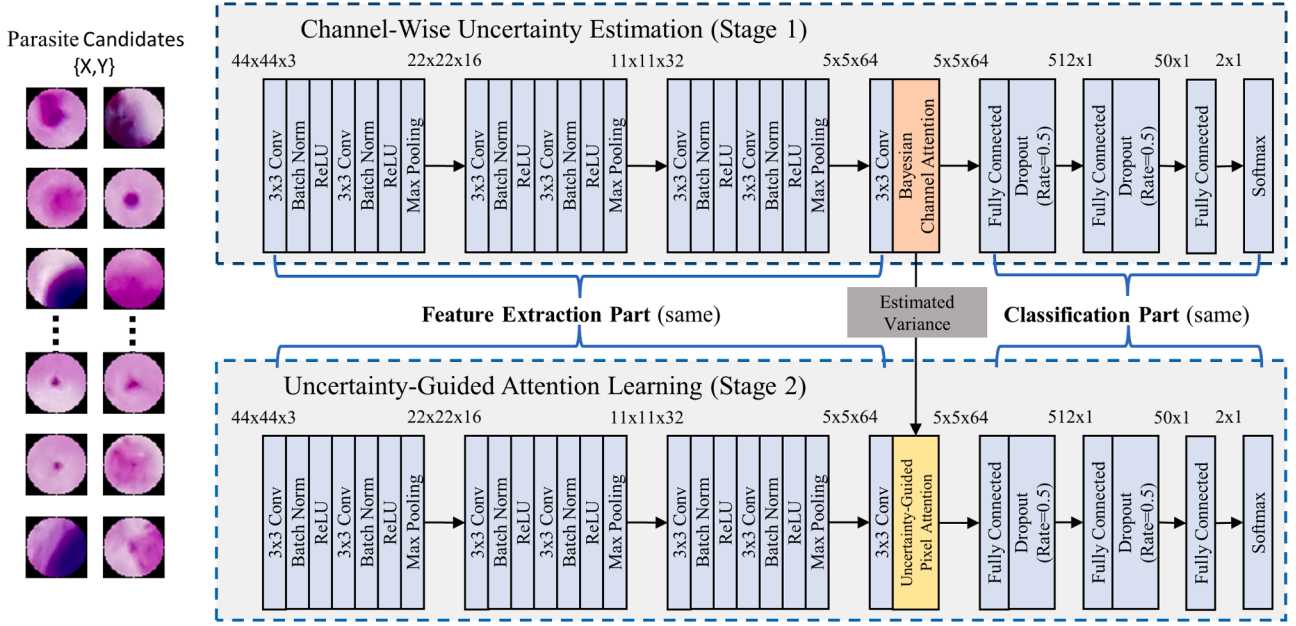
Fig. 3. Our method consists of two stages. Initially, Bayesian channel attention estimates the channel-wise uncertainty of the feature map (size $5 \times 5 \times 64$) produced by the last convolutional layer of Feature Extraction. Then, with the identified reliable channel features, we substitute Bayesian channel attention with uncertainty-guided pixel attention to extract informative pixel-wise features for the parasite identification.

channel attention and then illustrate how to incorporate it with uncertainty estimation.

3.4.1. Channel attention

In our network, the feature map $\mathbf{F} \in \mathbb{R}^{C \times H \times W}$ extracted by the last convolutional layer is utilized for parasite classification. We therefore study the uncertainties of its channels in the following sections. Here, $C = 64$, $H = 5$ and $W = 5$ denote the number of channels, height and width, respectively. In channel attention (Dai et al., 2021), we have:

$$\mathbf{g}_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{F}_c(i, j), \tag{1}$$

where $\mathbf{F}_c(i, j)$ denotes the value of cth channel $\mathbf{F}_c$ at position (i, j), and Eq. 1 is the global pooling function that change the shape of feature map $\mathbf{F}$ from $64 \times 5 \times 5$ to $64 \times 1 \times 1$.

Then, $\mathbf{g}_c$ will pass through two convolutional layers, and sigmoid, ReLU activation functions to obtain the attention weights of each channel:

$$\mathbf{a}_o = \sigma(Conv(\delta(Conv(\mathbf{g}_c)))), \tag{2}$$

where σ and δ are the sigmoid function and ReLU function, respectively. Besides, $\mathbf{a}_o$ refers to a vector of length $C = 64$ containing the attention weights for all channels in $\mathbf{F}$. With the sigmoid function σ, the values of $\mathbf{a}_o$ is normalized between 0 and 1.

Finally, we element-wise multiply the input $\mathbf{F}$ and the attention weights $\mathbf{a}_o$, and add them back to the input $\mathbf{F}$, generating $\hat{\mathbf{F}}$ utilized for downstream tasks:

$$\hat{\mathbf{F}} = \mathbf{F} + \mathbf{a}_o \otimes \mathbf{F}. \tag{3}$$

3.4.2. Uncertainty via Gaussian distribution

To consider uncertainty estimation into channel attention, we first introduce the variable $\mathbf{z}$ as attention score. In our case, $\mathbf{z}$ follows the *Gaussian* distribution which is simple and efficient to model uncertainty using variance.

$$p(\mathbf{z}|\mathbf{x}, \boldsymbol{\omega}) = \mathcal{N}(\mu(\mathbf{x}, \boldsymbol{\omega}), diag(\sigma^2(\mathbf{x}, \boldsymbol{\omega}))), \tag{4}$$

where $p(\boldsymbol{\omega}) = \mathcal{N}(0, \tau^{-1}\mathbf{I})$, and $\boldsymbol{\omega}$ refers to the neural network weights and has zero-mean isotropic Gaussian prior with precision τ. Meanwhile, $\mu(.)$

and $\sigma^2(.)$ are the mean and variance of the *Gaussian* distribution. It is worth noting that the variance $\sigma^2(\mathbf{x}, \boldsymbol{\omega})$ is dependent of input data $\mathbf{x}$ and thus can better express the amount of uncertainty from data. With attention score $\mathbf{z}$, the attention weights $\mathbf{a}_c$ can be obtained via:

$$\mathbf{a}_c = \sigma(\mathbf{z}). \tag{5}$$

As can be seen, the attention score $\mathbf{z}$ containing uncertainty passes through the sigmoid function, ensuring the generated attention weights $\mathbf{a}_c$ are within the range between 0 and 1. The detailed structure of our Bayesian channel attention is shown in Fig. 4 (a).

Now, suppose we have N input images $\mathbf{X} \in \mathbb{R}^{N \times P}$, let $\mathbf{Z} = [\mathbf{z}^{(1)}, .., \mathbf{z}^{(N)}]$ be the corresponding attention scores. Since we have N input images, there are N corresponding attention scores. Then, in the learning stage, the objective is to optimize the log-likelihood function:

$$\log p(\mathbf{Y}|\mathbf{X}) = \log \int \int p(\mathbf{Y}|\mathbf{Z}, \mathbf{X}, \boldsymbol{\omega})p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\omega})p(\boldsymbol{\omega})d\mathbf{Z}d\boldsymbol{\omega}. \tag{6}$$

3.5. Variational inference

The log-likelihood in Eq. 6 is intractable due to the non-linearity of the network. As per Gal and Ghahramani (2016), we obtain the evidence lower bound (ELBO) of $\log p(\mathbf{Y}|\mathbf{X})$ such that we are able to train the model on ELBO rather than on $\log p(\mathbf{Y}|\mathbf{X})$ due to its intractability. To obtain the ELBO of $\log p(\mathbf{Y}|\mathbf{X})$, we apply variational inference by introducing the following variational distributions:

$$q(\mathbf{Z}, \boldsymbol{\omega}|D) = q(\mathbf{Z}|\boldsymbol{\omega}, D)q(\boldsymbol{\omega}|D), \tag{7}$$

where $D = \{\mathbf{X}, \mathbf{Y}\}$. After introducing variational distribution $q(\mathbf{Z}, \boldsymbol{\omega}|D)$, we have:

$$\log p(\mathbf{Y}|\mathbf{X}) = \log q(\mathbf{Z}, \boldsymbol{\omega}|D)\frac{p(\mathbf{Y}|\mathbf{X})}{q(\mathbf{Z}, \boldsymbol{\omega}|D)}. \tag{8}$$

By applying Jensen's inequality, we obtain the ELBO of $p(\mathbf{Y}|\mathbf{X})$:

$$\log p(\mathbf{Y}|\mathbf{X}) \geq \int \int q(\mathbf{Z}, \boldsymbol{\omega}|D) \log \frac{p(\mathbf{Y}|\mathbf{X})}{q(\mathbf{Z}, \boldsymbol{\omega}|D)} d\mathbf{Z}d\boldsymbol{\omega}. \tag{9}$$

We further factorize the ELBO in Eq. 9 as following:

$$\begin{aligned} \log p(\mathbf{Y}|\mathbf{X}) \\ \geq \int \int q(\mathbf{Z}|\boldsymbol{\omega}, D)q(\boldsymbol{\omega}|D) \log p(\mathbf{Y}|\mathbf{Z}, \mathbf{X}, \boldsymbol{\omega})d\mathbf{Z}d\boldsymbol{\omega} \\ - KL[q(\mathbf{Z}|\boldsymbol{\omega}, D)||p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\omega})] - KL[q(\boldsymbol{\omega}|D)||p(\boldsymbol{\omega})]. \end{aligned} \tag{10}$$
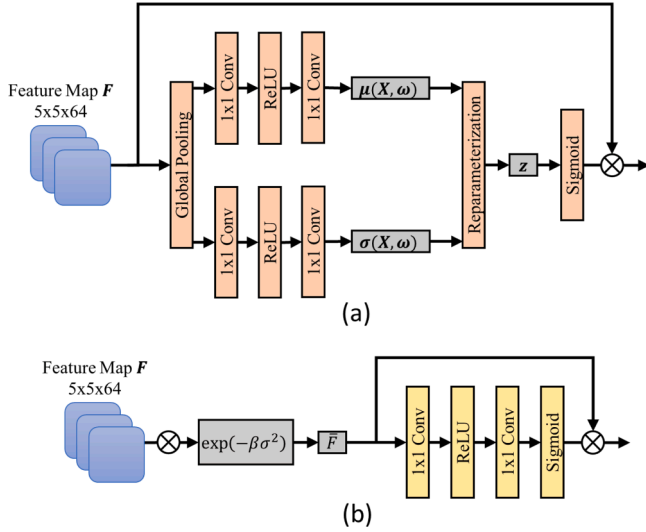
Fig. 4. The proposed attention mechanisms: (a) Bayesian Channel Attention - estimating the uncertainties of channels in the feature map to identify reliability of channel features. (b) Uncertainty-Guided Pixel Attention - exploiting the estimated uncertainties to filter out unreliable channel features which have larger uncertainty, making the network focused on the pixel-wise informative features from reliable channel features to facilitate more accurate classification of parasites.

Here, KL[·] denotes the KL-divergence between two distributions. The first KL term vanishes as two distributions are equivalent, and the second KL term becomes the ℓ_2 regularization of variational parameters in variational distribution $q(\mathbf{Z}, \boldsymbol{\omega}|D)$. In Eq. 10, the expectation is approximated using Monte Carlo sampling. First, we sample the weights $\tilde{\boldsymbol{\omega}} \sim q(\boldsymbol{\omega}, D)$ via Dropout as Bayesian Approximation (Gal & Ghahramani, 2016) with dropout masks. With sampled $\tilde{\boldsymbol{\omega}}$, the attention score $\mathbf{z}$ is subsequently sampled via reparameterization trick:

$$\tilde{\mathbf{z}} = \mu + \sigma\epsilon, \epsilon \sim N(0, \mathbf{I}). \tag{11}$$

Here, $\tilde{\mathbf{z}}$ refers to the sampled attention score. Then, Eq. 10 becomes:

$$\log p(\mathbf{Y}|\mathbf{X}) \geq \sum_{n=1}^{N} \log p(\mathbf{y}_n|\tilde{\mathbf{z}}_n, \mathbf{x}_n, \tilde{\boldsymbol{\omega}}) - ||\mathbf{Q}||^2 - ||\boldsymbol{\omega}||^2. \tag{12}$$

Here, $\mathbf{Q}$ is the variational parameter of $q(\mathbf{Z}|\boldsymbol{\omega}, D)$ defined in Eq. 7. The KL-divergence becomes the ℓ_2 regularization for the variational parameters. With sampled $\tilde{\mathbf{z}}_n$ and $\tilde{\boldsymbol{\omega}}$, Eq. 12 refers to the standard network training shown in Fig. 2.

3.6. Uncertainty-guided pixel attention

With estimated channel-wise uncertainty, we introduce the uncertainty-guided pixel attention. Traditional pixel attention (Qin et al., 2020) aims to extract important pixel-wise information from the feature map to facilitate downstream tasks. Unlike traditional pixel attention, our pixel attention learning leverages reliability weights estimated at the Bayesian channel attention stage to suppress unreliable channels. That is, important pixel-wise information is mainly extracted from reliable channel features that have a lower uncertainty.

For each channel, we assign it a weight that is determined by estimated uncertainty (variance σ^2). The weight indicates the reliability of its associated channel feature. Its value is normalized between 0 and 1 by applying exponential function to variance σ^2:

$$\mathbf{w}_\sigma = \exp(-\beta\sigma^2), \tag{13}$$

where $\mathbf{w}_\sigma$ is the reliability weight of size 64×1, and it is also visualized in Fig. 5. Besides, β is a hyperparameter with positive value to ensure $\mathbf{w}_\sigma$ is within $[0, 1]$. Then, we multiply feature $\mathbf{F}$ by $\mathbf{w}_\sigma$, such that
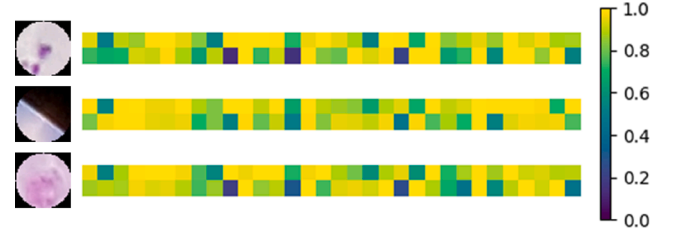

Fig. 5. Visualization of reliability weight w_σ. Left column: input image patches randomly selected from the Thick Smear 150 dataset. The top one is the positive sample containing malaria, while the bottom two patches are negative samples. Middle column: their corresponding reliability weights w_σ that appear as 64 bits, for which each bit indicates the uncertainty of one channel feature. Right column: color map of weight w_σ in which darker color indicates higher uncertainty, while the yellow one shows more reliability with less uncertainty. The channel features are generally reliable except for several channels with high uncertainties. We also observe that the estimated feature reliabilities are input-dependent due to varying input data noise levels.

$\tilde{\mathbf{F}} = \mathbf{F} \otimes \mathbf{w}_\sigma$. As a result, for a given channel from $\mathbf{F}$, if its corresponding value in $\mathbf{w}_\sigma$ is closer to 1, then the information from that channel will be maximally utilized in our pixel attention learning and vice versa.

With more reliable feature $\tilde{\mathbf{F}}$, we apply pixel attention to further extract effective pixel-wise information from it.

$$\mathbf{a}_p = \sigma(Conv(\delta(Conv(\tilde{\mathbf{F}})))), \tag{14}$$

where $\mathbf{a}_p$ is the pixel-wise attention weights of size $H \times W$ indicating pixel-wise importance on feature $\tilde{\mathbf{F}}$. Next, we apply element-wise multiplication for $\tilde{\mathbf{F}}$ and $\mathbf{a}_p$, and add back the multiplication output to $\tilde{\mathbf{F}}$ to generate the $\hat{\mathbf{F}}$ for our classification task.

$$\hat{\mathbf{F}} = \tilde{\mathbf{F}} + \mathbf{a}_p \otimes \tilde{\mathbf{F}}. \tag{15}$$

The derived feature $\hat{\mathbf{F}}$ contains important pixel-wise information that are more reliable with less uncertainty. As a result, it is able to further enhance the parasite classification accuracy of our model.

4. Experiments

4.1. Datasets and evaluation

To validate the robustness and generalizability of our method, we evaluated on two publicly available datasets that were independently collected by different medical centers. These datasets differ in terms of imaging equipment, sample preparation protocols, and staining conditions, which makes them suitable for a thorough performance evaluation.

4.1.1. Clinical malaria microscopy (CMM)

Thick blood films of 13 patients stained with Giemsa were collected by the University College Hospital (UCH) in Ibadan, Nigeria. Manescu et al. (2020b) UCH microscopists manually identified the malaria parasites in 100× magnification thick blood films, for which 2986 parasites were identified with an average size of 42×44 pixels. These thick blood films were then digitized to create 239 RGB images of resolution 2560×2160 for training and evaluation.

4.1.2. Thick smears 150

This dataset includes 150 patients, whose Giemsa-stained thick blood smear slides were photographed with 100× magnification at Chittagong Medical College Hospital, Bangladesh, using a smartphone camera from the different microscopic field of views. Yang et al. (2020) 1819 RGB images with the 3024×4032 resolution were generated. An expert reader from the Mahidol-Oxford Tropical Medicine Research Unit at Bangkok,

Table 1

Statistics of the Clinical Malaria Microscopy (CMM) and Thick Smears 150 datasets.

	CMM	Thick Smears 150
# of Subjects	13	150
# of Images	239	1,819
# of Parasites	2,986	84,961
Image Resolution	2560×2160	3024×4032
Avg Parasite Size	42×44	44
(in Pixels)	(Width x Height)	(Radius)

Thailand manually identified in these 84,961 malaria parasites (average radius of 22 pixels).

In Table 1, it shows a brief summary of above two datasets.

4.1.3. Evaluation metrics

We utilize the metrics of Precision, Recall, F1 score and average precision (AP):

$$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN},$$
$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, AP = \int_{r=0}^{1} p(r)dr, \quad (16)$$

where TP, FP and FN refer to the True Positive and False Positive detected parasites, and False Negative ground truth parasites, respectively. Meanwhile, AP refers to the area under the Precision Recall curve (PR curve). In a PR curve, given the recall score r, $p(r)$ is its corresponding precision score.

4.2. Implementation details

We implemented our model using PyTorch 1.7.1 on a computer with CUDA 11.1, one GPU device Tesla T4 with 15.109GB memory, 30GB RAM and 8 Intel (R) Xeon (R) Platinum 8259CL CPUs. The input to our model was image patches (parasite candidates) of size 44×44, and the batch size used was 2000. We evaluate each dataset using a 5-fold cross validation, with each fold including 4/5 and 1/5 of patients as the training and test sets, respectively. Out of the training set, 0.5% of the data were utilized as the validation set. Accordingly, the model with best performance on validation set was used to evaluate on the test set.

As introduced in Section 3, the optimization is executed in two stages that contain two networks. The first exploits Bayesian channel attention to estimate channel-wise feature map uncertainty, then the second utilizes the estimated uncertainties to guide the pixel attention learning.

To train both networks, we deployed stochastic gradient descent (SGD) with the learning rate of 0.0005 and momentum of 0.9. The total number of training epochs was 200 for both datasets. In the inference, we only utilized the second network for malaria parasites classification.

As per Yang et al. (2020), we select the parasite candidates before training our model. In our experiments, we selected 300 and 500 parasite candidates per image for the CMM and Thick Smears 150 datasets, respectively. Then, we generated 71,700 and 909,500 image patches as candidates for CMM and Thick Smears 150. Out of these, CMM had 94.52% of true positive parasites successfully identified, while for Thick Smears 150 we correctly detected 97.29% of parasites. Since CMM is much smaller than Thick Smears 150, it has fewer parasites than Thick Smears 150. Thus, we selected 300 candidates instead of 500, to reduce the number of potentially detected false positive parasites, which may degrade the model's performance.

4.3. Comparison with state-of-the-Art methods

We compare our method with six state-of-the-art baselines: Parasite Detection ConvNet (PDNet) (Yang et al., 2020), Modified Yolov4 (Yolov4-Mod) (Abdurahman et al., 2021), Yolov5 (Jocher, 2020), Yolov5opt (Sukumarran et al., 2024b), NesT (Zhang et al., 2022), RegionViT (Chen et al., 2022), and TransMIL (Shao et al., 2021). PDNet identifies parasites by performing image classification of the parasite candidates extracted from a blood smear image. Yolov5, Yolov5opt and Yolov4-Mod are object detection based methods, where Yolov5 detects more general objects in an image, Yolov5 is optimized as Yolov5opt by fine-tuning its training parameters to localize malaria-infected cells in thin blood smear images and Yolov4-Mod identifies malaria cells in a blood smear image by extending feature scales of Yolov4 and adding more detection layers to enhance its capability to detect small objects. Nest, RegionViT and TransMIL are transformer-based image classification methods. RegionViT aims to extract both global information and locally attended features, while Nest and TranMIL focus on long range information only.

We perform both parasite-level and patient-level evaluations. The former measures the performance on all the parasites in the entire dataset, while the latter evaluates the accuracy for each patient individually and averages the accuracy values across all the patients, to show the mean accuracy value with its standard deviation. As reported in Table 2, our method outperforms the baseline methods on both the CMM and Thick Smears 150 datasets. For CMM, the proposed method achieves the highest accuracy with respect to most metrics but has slightly inferior recall than that of Yang et al. (2020) for patient-level evaluation.

Table 2

Quantitative evaluation of our method and state-of-the-art methods on CMM and Thick Smear 150. We evaluate the parasite detection performance at both the parasite and patient levels. For patient-level evaluation, we utilize mean value $\pm$ standard deviation to show variance across patients. The best result is highlighted in boldface.

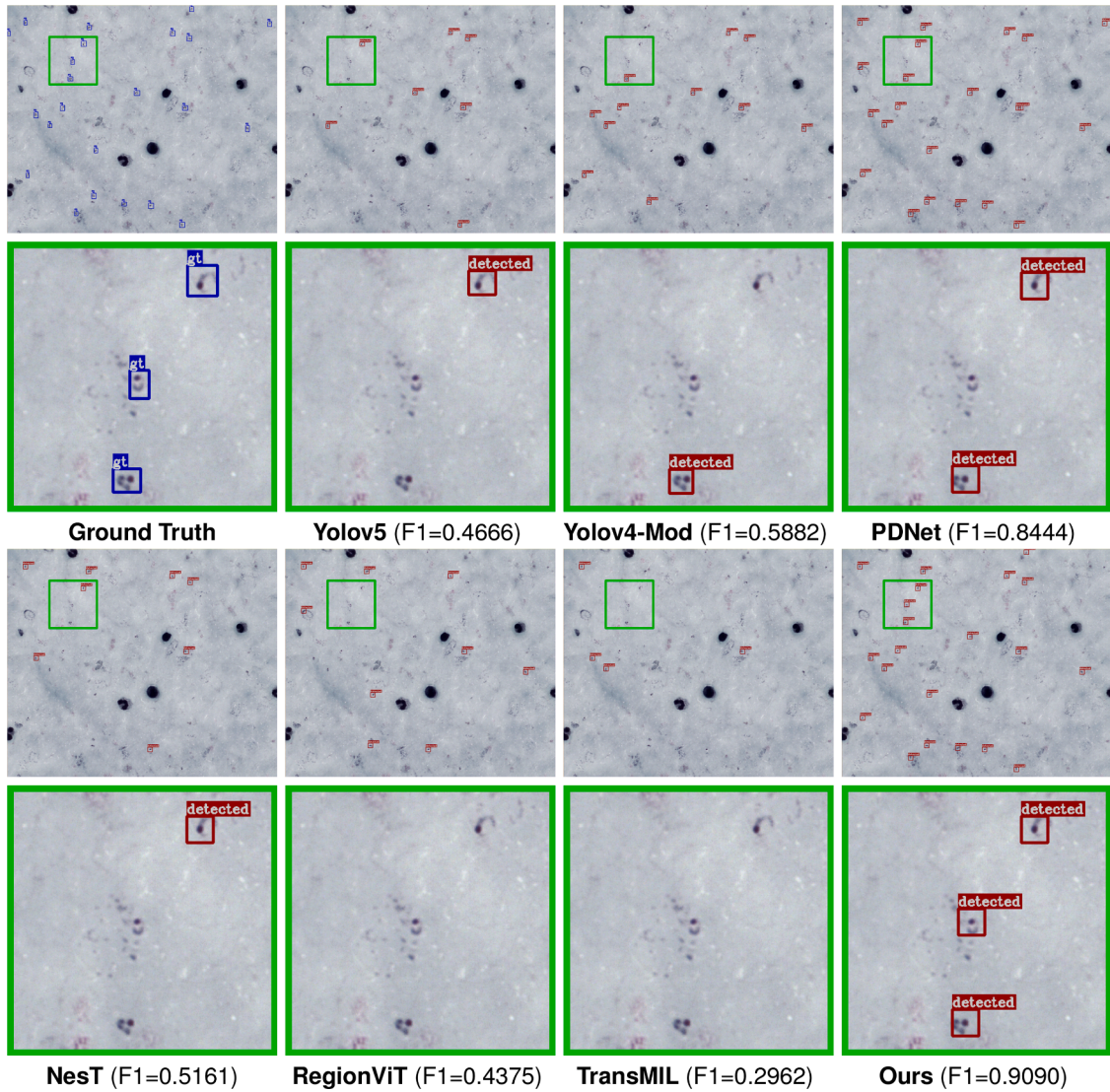| Methods	Dataset	Parasite-Level Evaluation				Patient-Level Evaluation			
		Precision	Recall	F1 Score	AP	Precision	Recall	F1 Score	AP
Yolov5 (Jocher, 2020)	CMM	49.07	63.44	55.34	37.48	23.51±27.23	42.61±32.12	27.45±26.05	25.66±21.97
Yolov5opt (Sukumarran et al., 2024b)	CMM	29.67	68.59	41.42	23.52	16.50±18.74	37.08±28.95	20.85±20.93	15.16±14.99
Yolov4-Mod (Abdurahman et al., 2021)	CMM	45.72	48.36	47.00	25.86	24.92±25.09	39.03±36.36	26.40±24.57	21.48±20.02
PDNet (Yang et al., 2020)	CMM	46.18	67.85	54.96	30.31	29.88±17.40	**58.34**±33.46	34.21±19.39	33.04±18.13
NesT (Zhang et al., 2022)	CMM	41.03	21.70	28.39	16.56	34.61±18.98	22.57±22.97	22.49±13.37	16.90±12.05
RegionViT (Chen et al., 2022)	CMM	43.44	22.18	29.36	14.72	30.33±15.41	25.34±23.60	23.49±13.96	12.73±11.39
TransMIL (Shao et al., 2021)	CMM	32.30	22.92	26.81	11.94	31.61±16.97	53.62±25.95	33.38±17.51	23.16±15.41
ours	CMM	**51.39**	**71.96**	**59.96**	**40.73**	**35.22**±18.26	56.62±33.86	**37.40**±20.24	**36.39**±20.81
Yolov5 (Jocher, 2020)	Thick Smear 150	78.99	83.52	81.19	73.83	69.24±20.32	72.78±19.00	69.46±18.84	61.26±19.83
Yolov5opt (Sukumarran et al., 2024b)	Thick Smear 150	75.84	67.41	71.38	51.67	64.74±20.31	61.56±15.67	61.64±16.98	46.49±17.29
Yolov4-Mod (Abdurahman et al., 2021)	Thick Smear 150	77.99	81.27	79.60	72.22	69.26±18.72	71.98±19.12	68.77±17.75	60.50±19.69
PDNet (Yang et al., 2020)	Thick Smear 150	82.19	82.74	82.46	76.63	69.21±24.86	79.51±15.21	70.91±20.07	67.82±18.75
NesT (Zhang et al., 2022)	Thick Smear 150	76.82	80.27	78.51	75.99	65.71±28.04	76.75±17.21	66.76±22.99	63.97±20.73
RegionViT (Chen et al., 2022)	Thick Smear 150	78.31	81.40	79.82	75.92	64.48±26.28	79.98±15.26	67.57±21.65	66.53±19.76
TransMIL (Shao et al., 2021)	Thick Smear 150	80.64	82.17	81.66	76.26	66.38±25.46	79.57±15.49	69.03±20.63	66.59±19.28
ours	Thick Smear 150	**83.15**	**85.13**	**84.13**	**77.77**	**69.35**±23.73	**81.58**±14.34	**72.31**±19.02	**69.65**±17.83

Fig. 6. Malaria parasites detection examples in CMM. 1st and 3rd rows show the raw thick blood smears with the detected parasites, while 2nd and 4th rows – the associated regions of interest (ROI) within the green bounding boxes. Boxes marked 'gt' and 'detected' denote the ground truth and detected parasites, and we show the F1 score for each example.

For Thick Smear 150, our method clearly outperforms all baselines. Besides that, the performance of all the methods on the CMM dataset is inferior to that on Thick Smears 150. This is presumably due to the fact that the former offers fewer training samples than the latter.

It is also worth noting that the classification based methods, such as Yang et al. (2020) and ours, are generally better than the object detection based methods. This is primarily due to the fact that the parasites are tiny objects in high resolution thick blood smear images, which renders the object detection methods harder to identify them. By contrast, our method pre-selects a large number of parasite candidates that include most parasites and then classifies these candidates as parasites or non-parasites. Another reason is that more training samples helps to enhance the model performance. Specifically, object detection methods treat individual images as training samples, while our method utilizes the pre-selected image patches (parasite candidates) for training. That said, our method has more training samples than the baselines, especially for a small scale dataset such as CMM. By further estimating the channel-wise uncertainties caused by data noises, our method is able to make predictions relying on reliable features and thus achieves the best performance. In Figs. 6 and 7, we demonstrate the visual comparisons between all methods on malaria parasite detection. Since the detected

parasites are small objects in thick blood smears, we also show a cropped ROI area from the original output.

In Table 2 we also observe that modern transformer-based methods do not demonstrate superior performance, and their accuracy is in general lower than that of CNN-based methods such as PDNet and ours. This can presumably be attributed to overfitting that occurs when training on these datasets. Normally, advanced transformers involve a very large number of parameters that are learned from large datasets. However, the available blood smear datasets of malaria parasites are scarce, as the data is hard to collect. Hence, the number of training samples in our case cannot sufficiently optimize the transformers with many parameters. It can be seen that transformers trained on the CMM dataset (contains 239 images) achieve substantially lower accuracy levels than those trained on Thick Smear 150 (1819 images). Another thing to highlight is that our input images preselected for parasite classification are cropped as size 44×44, which is likely to also cause transformers overfit when training on such small sized images. In fact, CNN-based networks with several convolutions would be sufficient for extraction of long range information on such small sized images. Additionally, we also attempted several object detection based transformers to detect parasites directly from the whole blood smear image. However, we failed to train these
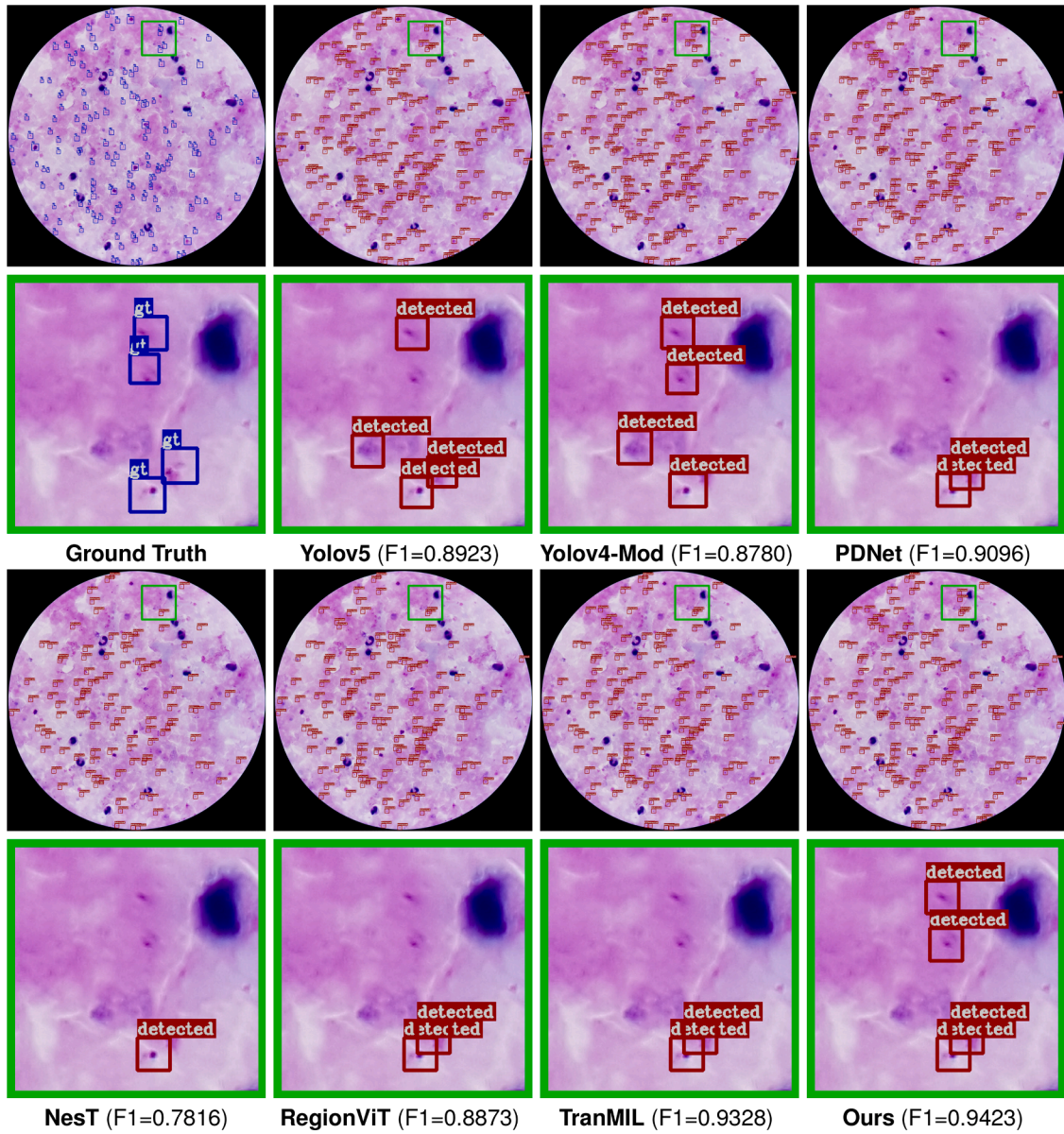
Fig. 7. Malaria parasites detection examples of ground truth and comparison methods on Thick Smear 150 dataset. Note that 1st and 3rd rows show the raw thick blood smears with the detected parasites and the remaining rows – the associated regions of interest (ROI) within green bounding boxes. Here, the bounding box marked 'gt' and 'detected' refers to the ground truths and detected parasites, and we also show the F1 score for each example.

transformers successfully. The available training samples were insufficient to train the transformer models, especially the models having a large number of parameters to optimize. Also, the parasites are tiny objects that makes them more difficult to detect in high-resolution blood smears.

4.4. Computational analysis

To evaluate the practicality of our proposed two-stage framework in real-world applications, we conduct a comprehensive computational analysis and compare it against aforementioned baselines.

As shown in Table 4, our method exhibits competitive efficiency in terms of both computational load and inference time. Specifically, the overall number of parameters is only 1.0M. In terms of FLOPs, we observe that our model respectively requires 6.2G for CMM and 10.5G for thick smear 150. These values are substantially lower than those of many other methods. It is worth noting that the FLOPs of detection-based models such as YOLOv4-Mod, YOLOv5, and YOLOv5opt remain

identical across the two datasets (CMM and Thick Smear 150). This is because these models operate on standard resized inputs (640×640), leading to a constant computational load regardless of dataset-level variations. In contrast, classification-based frameworks, including ours and other patch-level models such as PDNet, show different FLOPs across the two datasets. This is due to the fact that these models perform classification on candidate patches, and the number of selected patches differs across datasets. In our pipeline, each candidate patch is processed independently through the classification network, so the total FLOPs scale linearly with the number of patches. For example, the average number of candidate patches is 300 per image in the CMM dataset, and 500 in Thick Smear 150. As a result, although the network architecture remains unchanged, the total FLOPs increase with the number of patches.

In addition, our inference time remains 0.821s for CMM and 2.146s for Thick Smear 150, indicating that our method is deployable in high-throughput screening settings for large-scale parasite detection. Although our approach operates in a two-stage manner, it includes

Table 3
Effectiveness of applying different attention mechanisms for parasite detection in the CMM and Thick Smear 150 datasets. For patient-level evaluation, we utilize mean value ± standard deviation to show variance across patients. The best result is highlighted in boldface.

Methods	Dataset	Parasite-Level Evaluation				Patient-Level Evaluation			
		Precision	Recall	F1 Score	AP	Precision	Recall	F1 Score	AP
self attn	CMM	45.01	65.73	53.43	28.74	23.70±23.75	53.19±31.02	29.43±16.49	22.80±13.82
Bayes ch attn	CMM	49.06	65.12	55.96	35.87	33.94±19.23	53.03±33.34	34.90±20.14	30.56±17.61
ch attn	CMM	49.46	68.42	57.42	33.34	**38.70**±19.61	54.67±33.53	**38.64**±19.87	32.37±18.11
pxl attn	CMM	51.06	71.55	59.60	39.98	34.75±18.20	56.57±33.83	37.03±20.19	35.39±21.03
multi-head attn	CMM								
– 1-head		48.55	67.65	56.53	36.07	35.97±19.28	53.45±30.44	37.12±20.24	34.64±18.35
– 2-head		43.78	64.86	52.27	33.13	31.49±19.76	55.59±31.64	34.96±21.19	31.86±16.08
– 4-head		41.80	56.04	47.88	28.40	28.96±17.93	55.12±33.54	32.81±20.20	31.60±19.43
ours	CMM	**51.39**	**71.96**	**59.96**	**40.73**	35.22±18.26	**56.62**±33.86	37.40±20.24	**36.39**±20.81
self attn	Thick Smear 150	81.05	81.57	81.31	76.51	68.08±23.99	79.41±19.79	69.36±20.43	68.63±19.97
Bayes ch attn	Thick Smear 150	82.50	83.71	83.10	76.56	68.61±24.05	80.91±13.87	71.38±19.24	69.13±18.21
ch attn	Thick Smear 150	81.90	84.54	83.20	76.28	67.88±24.76	81.07±14.51	70.95±19.83	68.96±18.57
pxl attn	Thick Smear 150	82.60	83.94	83.26	76.49	68.47±24.30	80.75±14.22	71.18±19.31	68.74±18.31
multi-head attn	Thick Smear 150								
– 1-head		82.43	83.47	82.95	76.65	68.55±24.44	80.16±15.13	70.87±19.72	68.30±18.67
– 2-head		79.67	83.26	81.43	74.89	66.20±25.18	78.97±18.43	67.95±20.95	66.11±20.52
– 4-head		77.75	84.28	80.88	74.96	63.13±25.50	81.32±14.56	67.71±21.09	67.07±19.54
ours	Thick Smear 150	**83.15**	**85.13**	**84.13**	**77.77**	**69.35**±23.73	**81.58**±14.34	**72.31**±19.02	**69.65**±17.83

Table 4
Comparison of computational cost across different baseline methods. We illustrate the number of parameters (in millions), FLOPs (in Giga), and inference time (in seconds per image) across both CMM and thick smear 150 datasets.

Methods	Params (M)	FLOPs (G)		Inference Time (s)	
		CMM	Thick Smear 150	CMM	Thick Smear 150
Yolov5 (Jocher, 2020)	9.1	12.1	12.1	0.007	0.007
Yolov5opt (Sukumarran et al., 2024b)	25.1	32.3	32.3	0.014	0.014
Yolov4-Mod (Abdurahman et al., 2021)	46.4	74.4	74.4	0.022	0.022
PDNet (Yang et al., 2020)	0.9	6.2	10.4	0.818	2.148
NesT (Zhang et al., 2022)	16.1	8.1	13.5	0.818	2.158
RegionViT (Chen et al., 2022)	12.5	36.6	61.1	0.819	2.171
TransMIL (Shao et al., 2021)	2.1	1.2	2.1	0.818	2.068
Ours	1.0	6.2	10.5	0.821	2.146

the lightweight backbone specifically tailored for the small patch size (44 × 44) and effectively reduces the computation required during classification, providing rapid inference while maintaining strong classification performance. We also observe that object detection methods such as YOLOv5 have faster inference time than classification based methods like ours. This is primarily because the inference time of classification based methods also involves parasites candidates preselection. Therefore, these results demonstrate that our model achieves a favorable trade-off between accuracy and efficiency, validating its suitability for practical deployment in malaria parasite screening.
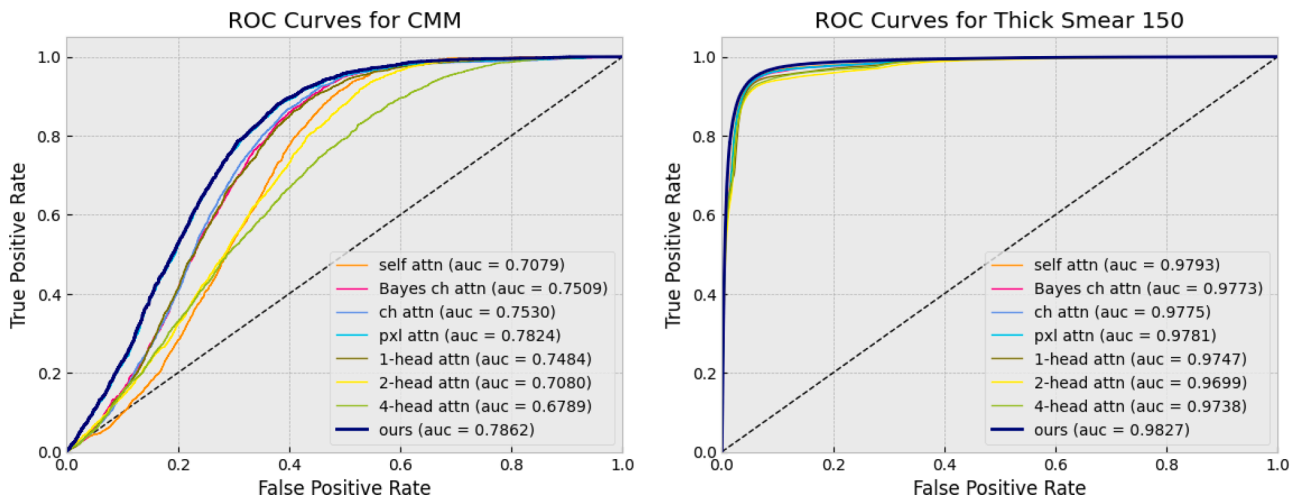

Fig. 8. ROC curves of different attention mechanisms for true and false parasites classification of pre-selected candidates in the CMM and Thick Smear 150 datasets. Here, the multi-head attention with 1 head, 2 heads and 4 heads are respectively denoted as 1-head attn, 2-head attn and 4-head attn.

Table 5
Sensitivity of hyperprameter β on CMM and Thick Smear 150. We set $\beta = 10$ as the optimal value and the best result is highlighted in boldface.

Hyperparameter	Dataset	F1 Score	AP
$\beta = 1$	CMM	59.82	39.80
$\beta = 10$	CMM	**59.96**	**40.73**
$\beta = 20$	CMM	59.87	40.56
$\beta = 100$	CMM	57.98	37.40
$\beta = 1$	Thick Smear 150	83.09	76.59
$\beta = 10$	Thick Smear 150	**84.13**	**77.77**
$\beta = 20$	Thick Smear 150	82.51	77.03
$\beta = 100$	Thick Smear 150	81.22	68.20

4.5. Ablation study

Note that our method exploits (Yang et al., 2020) as the backbone, integrated with the uncertainty-guided attention learning. As shown in Table 2, our method outperforms the baseline (Yang et al., 2020) on both the datasets. Hence, we investigate in this ablation study how different attention mechanisms affect the parasite detection performance. We evaluate five typical attention mechanisms: (i) self attention (self attn) (Wang et al., 2018), (ii) channel attention (ch attn) (Dai et al., 2021), (iii) Bayesian channel attention (Bayes ch attn), (iv) pixel attention (pxl attn) Qin et al. (2020), and (v) multi-head attention (multi-head attn) (Vaswani et al., 2017). Self attention is extensively used to extract global information within the feature map. Channel attention estimates the importance of feature map channels, such that channels with higher estimated importance are utilized for downstream tasks. Bayesian channel attention is introduced in section 3.4, and it is utilized to estimate uncertainty for channels of the feature map. Pixel attention implies that different pixels in the feature map may have different information and assigns higher attention weights to the important pixels. Lastly, the multi-head attention jointly attends to information from different representation subspaces at different positions.

We illustrate the comparison of parasite prediction accuracy in Table 3. It is clear that our uncertainty-guided pixel attention outperforms the other methods. This can be attributed to the following reasons: 1) our attention mechanism involves channel-wise uncertainty estimation helping to identify more reliable channel features for subsequent tasks; 2) since pixel-wise features are inherently more flexible and effective than channel-wise features, we further extract important pixel-wise features based on these identified reliable channel features.

Recall that our method identifies parasites from a large number of pre-selected candidates. As the candidates may include false parasites, this may affect the performance of identifying true parasites. Hence, we evaluate the capability of our model to classify true parasites and false parasites within pre-selected candidates when equipped with the various attention mechanisms. We show the obtained ROC curves in Fig. 8 and the AUC of the proposed attention is the highest.

4.6. Hyperparameter sensitivity study

In our model, we exploit one hyperparameter β defined in Eq. 13. In the experiments, we set $\beta = 10$ based on the parasite level F1 score and AP. First, β is greater than or equal to 0 since the value of w_σ in Eq. 13 should be between 0 and 1. When $\beta = 0$, the value of w_σ is 1, which means that the effect of uncertainty is no longer considered. In that case, our uncertainty-guided pixel attention is equivalent to the traditional pixel attention. When $\beta \to +\infty$, the value of w_σ is close to 0, which limits the information propagation from all the channel features to our attention learning and renders the learning ineffective. In Fig. 9, we illustrate the effects of different β values on the generated reliability weights. We present two example input patches in Fig. 9 (a) and their channel-wise reliability weights produced by different β values ($\beta = 1, 10, 20, 100$) are shown in Fig. 9 (b). As seen in Fig. 9 (b), when
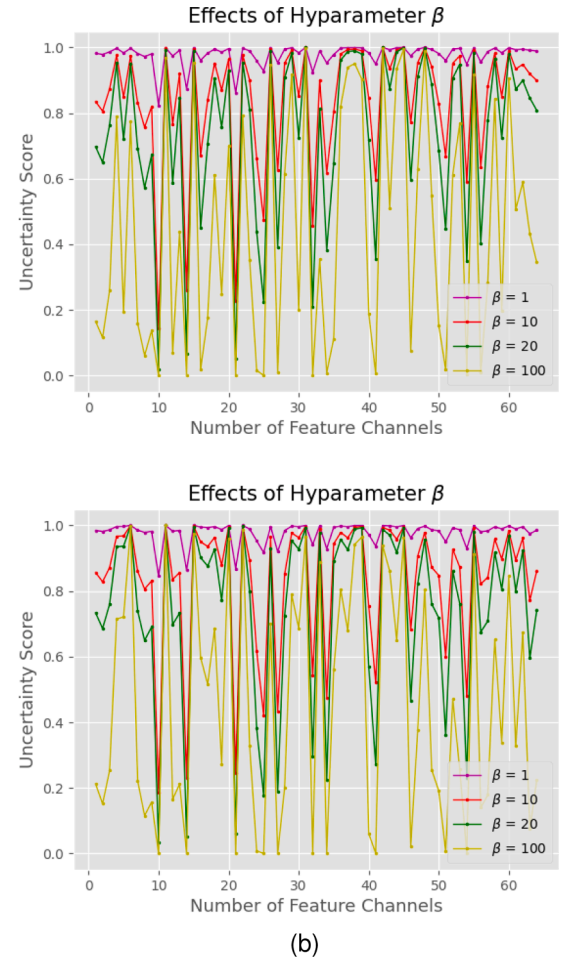

Fig. 9. Illustrations of the effects of different β values on generated reliability weights. (a) Two example input patches. (b) Reliability weights visualized for (a) under four different β values.

$\beta = 1$ (purple line) most reliability weights of these two examples are above 0.85, which means these scores are unable to filter out unreliable features. In contrast, when $\beta = 100$ (yellow line), many reliability weights are below 0.5, indicating that most features would be considered unreliable. These points are also illustrated in Table 5, where larger values of β yield poorer results. This suggests that higher β values hinder information propagation and consequently our pixel attention does not learn much. Likewise, a smaller β value like $\beta = 1$ yields lower F1 and AP, since many unreliable features are not excluded. Due to the high computational costs of training, we did not conduct an exhaustive search over β. Instead, we explored a reasonable range of w_σ and selected [1, 10, 20, 100] to demonstrate the effect of β on the reliability weights and model performance.

4.7. Effect of input patch size

In our work, we first identify a large number of parasite candidates and then crop them into patches of size $44 \times 44 \times 3$ for parasite detection. As per Yang et al. (2020), the patch size is determined by the average size of parasites within the dataset. In this section, we study how different patch sizes affect parasite detection performance. Other than the utilized size $44 \times 44 \times 3$, we also evaluate other sizes, including $36 \times 36 \times 3$, $52 \times 52 \times 3$ and $60 \times 60 \times 3$, on both CMM and Thick Smear 150 datasets.

In Fig. 10, we illustrate AUC score and both Parasite-Level and Patient-Level evaluations of Precision, Recall, F1 Score, AP for
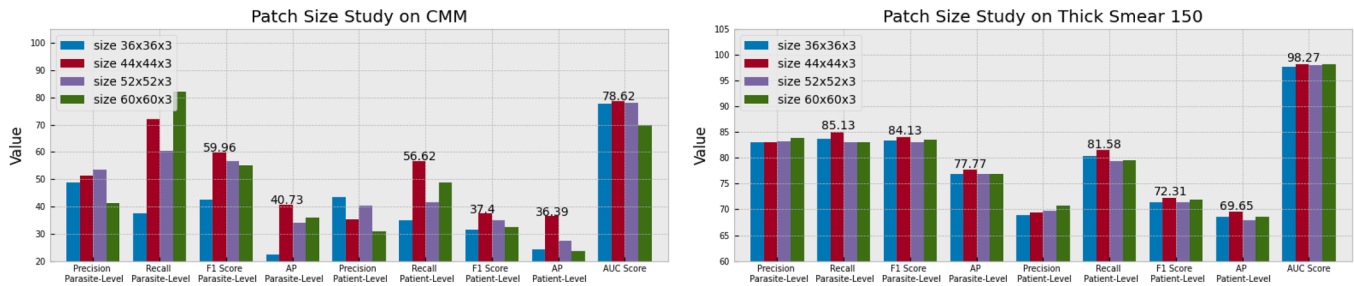
Fig. 10. The effects of different patch sizes on the CMM and Thick Smear 150 datasets. Out of all evaluation metrics, we indicate the best performance on top of red bars (size $44 \times 44 \times 3$). Clearly, the red bars with size $44 \times 44 \times 3$ outperform the remaining bars with respect to most metrics.

comparisons among different patch sizes. To better distinguish our performances from other patch sizes, we only present in Fig. 10 the scores of our utilized patch size for those best-performing metrics. We can see that the utilized size $44 \times 44 \times 3$ achieves 6 highest scores out of 9 metrics on CMM dataset, and has 7 best-performing metrics on Thick Smears 150 dataset. In general, the performance of patch size $44 \times 44 \times 3$ surpasses the rest. This is presumably because the patch size $36 \times 36 \times 3$ is too small to provide sufficient information for parasite detection, while the larger sizes $52 \times 52 \times 3$ and $60 \times 60 \times 3$ introduce more background noises that disturb the identification of parasites.

5. Discussion and conclusion

In this work, we present an uncertainty-guided attention learning for malaria parasites detection in thick blood smear images. By reformulating traditional channel attention under the Bayesian framework, we are able to estimate channel-wise uncertainty of the learned feature map. Subsequently, our model exploits pixel attention to extract important and informative pixel-wise features to facilitate parasite classification. With the estimated uncertainties, our pixel attention is able to extract pixel-wise information from more reliable channel features that have less uncertainty. Experiments using two public datasets show the effectiveness and superiority of our method over several state-of-the-art malaria parasite detection methods. Thus, our work offers a sound alternative to manual detection methods, allowing to estimate the malaria severity and prioritize patient treatment, which can help in under-developed malaria-endemic areas.

One of our work's most notable strengths lies in its methodological innovation. Unlike conventional attention mechanisms that focus on global contextual features or treat channels equally, our method explicitly quantify channel-wise uncertainty using estimated variances under a Bayesian formulation. This allows us to identify and emphasize reliable features, while suppressing noisy or unreliable ones. Such a formulation is particularly effective for tiny parasites in thick smear images, for which these tiny parasites are often obscured by background artifacts or staining irregularities. Moreover, our hybrid attention design, that combines Bayesian uncertainty modeling with pixel-level attention, is novel in the field, and has shown clear advantages in parasite classification.

Our method also provides interpretability that is particularly beneficial for high-stakes medical applications. The estimated uncertainty scores offer insight into the reliability of the input data. For instance, input patches dominated by unreliable features can be flagged as low-confidence predictions. This interpretability can assist healthcare professionals in deciding when to rely on automated predictions and when to need manual review.

In addition, our framework also exhibits strong potential for generalization. While this study focuses on malaria parasite detection in small image patches (44×44), the core structure is a binary classification network with explicit noise-handling capability. Since various types of medical imaging data inevitably suffer from acquisition noise, background artifacts, or sensing errors, our framework could be adapted to other classification tasks, such as histopathology or dermatology image

analysis. Some architectural adjustments, such as increasing the number of convolutional layers, may be required to accommodate larger images, but the fundamental idea of uncertainty-aware feature selection remains broadly applicable.

Another important strength of our method is its practical deployability. Although our network operates in two stages, it remains lightweight in terms of model size and FLOPs, with inference times under 2.2 seconds per image. Given that malaria diagnosis is a non-real-time screening task where accuracy and reliability are prioritized over latency, our model maintains an effective balance between performance and efficiency, making it suitable for use in clinical decision support systems or automated screening tools.

Despite that, our model has limitations. First, our performance partially depends on the quality of the parasite candidate preselection stage. If true parasites are not successfully selected in this stage, they cannot be recovered in subsequent classification, potentially leading to false negatives. Secondly, our method requires two-stage training, which may complicates integration into fully end-to-end pipelines. Nevertheless, our work introduces a robust and interpretable framework for malaria parasite detection, which not only achieves state-of-the-art performance but also offers practical value in clinical contexts. We hope future studies will explore theoretical analyses of how uncertainty-guided feature selection impacts model robustness and generalization, providing a more principled understanding of its advantages. Meanwhile, we believe our proposed idea that incorporating uncertainty modeling into attention mechanisms can inspire future advancements in reliable and trustworthy medical AI systems.

CRediT authorship contribution statement

Hao Xiong: Formal analysis, Methodology, Writing – original draft, Conceptualization; **Zhiyong Wang:** Supervision, Writing – review & editing; **Roneel V. Sharan:** Writing – review & editing; **Shlomo Berkovsky:** Supervision, Writing – review & editing.

Data availability

Data will be made available on request.

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.

References

Abdar, M., Samami, M., Dehghani Mahmoodabad, S., Doan, T., Mazoure, B., Hashemifesharaki, R., Liu, L., Khosravi, A., Rajendra Acharya, U., Makarenkov, V., & Nahavandi, S. (2021). Uncertainty quantification in skin cancer classification using three-way decision-based bayesian deep learning. *Computers in Biology and Medicine, 135*, 104418.

Abdul Nasir, A. S., Mashor, M. Y., & Mohamed, Z. (2012). Segmentation based approach for detection of malaria parasites using moving k-means clustering. In *IEEE-EMBS conference on biomedical engineering and sciences* (pp. 653–658).

Abdurahman, F., Fante, K. A., & Aliy, M. (2021). Malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models. *BMC Bioinformatics*, *22*(112).

Amersfoort, J. V., Smith, L., Teh, Y. W., & Gal, Y. (2020). Uncertainty estimation using a single deep deterministic neural network. In *International conference on machine learning* (pp. 9690–9700).

Amini, A., Schwarting, W., Soleimany, A., & Rus, D. (2020). Deep evidential regression. In *Advances in neural information processing systems* (pp. 14927–14937). (*vol. 30*).

Ashukha, A., Lyzhov, A., Molchanov, D., & Vetrov, D. (2019). Pitfalls of in-domain uncertainty estimation and ensembling in deep learning. In *International conference on learning representations*.

Baumgartner, C. F., Tezcan, K. C., Chaitanya, K., Hötker, A. M., Muehlematter, U. J., Schawkat, K., Becker, A. S., Donati, O., & Konukoglu, E. (2019). Phiseg: Capturing uncertainty in medical image segmentation. In *International conference on medical image computing and computer assisted intervention* (p. 119–127).

Cao, X., Chen, H., Li, Y., Peng, Y., Wang, S., & Cheng, L. (2020). Uncertainty aware temporal ensembling model for semi-supervised abus mass segmentation. *IEEE Transactions on Medical Imaging*, *40*(1), 431–443.

Chakrabortya, K., Chattopadhyayb, A., Chakrabarti, A., Acharyad, T., & Dasguptae, A. K. (2015). A combined algorithm for malaria detection from thick smear blood slides. *Journal of Health and Medical Informatics*, *6*, 1–6.

Chen, C.-F., Panda, R., & Fan, Q. (2022). Regionvit: Regional-to-local attention for vision transformers. In *International conference on learning representations*.

Dai, J., Li, Y., He, K., & Sun, J. (2016). R-FCN: Object detection via region-based fully convolutional networks. In *Advances neural information processing system* (pp. 379–387).

Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., & Barnard, K. (2021). Attentional feature fusion. In *Ieee winter conference on applications of computer vision* (pp. 3559–3568).

Dave, I. R., & Upla, K. P. (2017). Computer aided diagnosis of malaria disease for thin and thick blood smear microscopic images. In *International conference on signal processing and integrated networks* (pp. 561–565).

Delahunt, C. B., Mehanian, C., Hu, L., McGuire, S. K., Champlin, C. R., Horning, M. P., Wilson, B. K., & Thompon, C. M. (2015). Automated microscopy and machine learning for expert-level malaria field diagnosis. In *IEEE global humanitarian technology conference* (pp. 393–399).

Dev, A., Fouda, M. M., Kerby, L., & Md Fadlullah, Z. (2024). Advancing malaria identification from microscopic blood smears using hybrid deep learning frameworks. *IEEE Access*, *12*, 71705–71715.

Doering, E., Pukropski, A., Krumnack, U., & Schaffand, A. (2020). Automatic detection and counting of malaria parasite-infected blood cells. In *Medical imaging and computer-aided diagnosis* (pp. 145–157).

Elter, M., Hasslmeyer, E., & Zerfass, T. (2011a). Detection of malaria parasites in thick blood films. In *International conference of the IEEE engineering in medicine and biology society* (pp. 5140–5144).

Elter, M., Haßlmeyer, E., & Zerfaß, T. (2011b). Detection of malaria parasites in thick blood films. In *International conference of the IEEE engineering in medicine and biology society* (pp. 5140–5144).

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International conference on machine learning* (pp. 1050–1059).

Gantenbein, M., Erdil, E., & Konukoglu, E. (2020). RevPHiseg: A memory-efficient neural network for uncertainty quantification in medical image segmentation. In *Uncertainty for safe utilization of machine learning in medical imaging, and graphs in biomedical image analysis* (pp. 13–22).

Girshick, R. (2015). Fast r-CNN. In *IEEE international conference on computer vision* (pp. 1440–1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE conference on computer vision and pattern recognition* (pp. 1–8).

Gopakumar, G. P., Swetha, M., Siva, G. S., & Subrahmanyam, G. R. K. S. (2018). Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. *Journal of Biophotonics*, *11*(3), e201700003.

Guemas, E., Routier, B., Ghelfenstein-Ferreira, T., Cordier, C., Hartuis, S., Marion, B., Bertout, S., Varlet-Marie, E., Costa, D., & Pasquier, G. (2024). Automatic patient-level recognition of four plasmodium species on thin blood smear by a real-time detection transformer (RT-DETR) object detection algorithm: A proof-of-concept and evaluation. *Microbiology Spectrum*, *12*, e01440–23.

Hanif, N. S. M. M., Mashor, M. Y., & Mohamed, Z. (2011). Image enhancement and segmentation using dark stretching technique for plasmodium falciparum for thick blood smear. In *International colloquium on signal processing and its applications* (pp. 257–260).

Herzog, L., Murina, E., Dürr, O., Wegener, S., & Sick, B. (2020). Integrating uncertainty in deep neural networks for MRI based stroke analysis. *Medical Image Analysis*, *65*, 101790.

Hu, S., Worrall, D., Knegt, S., Veeling, B., Huisman, H., & Welling, M. (2019). Supervised uncertainty quantification for segmentation with multiple annotations. In *International conference on medical image computing and computer assisted intervention* (pp. 137–145).

Huang, L., Denoeux, T., Vera, P., & Ruan, S. (2022). Evidence fusion with contextual discounting for multi-modality medical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 401–411).

Jocher, G. (2020). ultralytics/yolov5: v3.1 – bug fixes and performance improvements. [Online]. Available: https://doi.org/10.5281/zenodo.4154370.

Kaewkamnerd, S., Intarapanich, A., Pannarat, M., Chaotheing, S., Uthaipibull, C., & Tongsima, S. (2011). Detection and classification device for malaria parasites in thick-blood films. In *International conference on intelligent data acquisition and advanced computing systems* (pp. 435–438). (*vol. 1*).

Kohl, S. A. A., Romera-Paredes, B., Meyer, C., Fauw, J. D., Ledsam, J. R., Maier-Hein, K. H., Eslami, S. M. A., Rezende, D. J., & Ronneberger, O. (2018). A probabilistic u-net for segmentation of ambiguous images. In *International conference on neural information processing systems* (p. 6965–6975).

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in neural information processing systems* (pp. 6405–6416). (*vol. 30*).

Leibig, C., Allken, V., Ayhan, M. S., Berens, P., & Wahl, S. (2017). Leveraging uncertainty information from deep neural networks for disease detection. *Scitific Reports*, *7*(1), 17816.

Lin, C., Wu, H., Wen, Z., & Qin, J. (2021). Automated malaria cells detection from blood smears under severe class imbalance via importance-aware balanced group softmax. In *International conference on medical image computing and computer assisted intervention* (pp. 455–465).

Loddo, A., Di Ruberto, C., & Kocher, M. (2018). Recent advances of malaria parasites detection systems based on mathematical morphology. *Sensors*, *18*(2), 513.

Madhu, G., Mohamed, A. W., Kautish, S., Shah, M. A., & Ali, I. (2023). Intelligent diagnostic model for malaria parasite detection and classification using imperative inception-based capsule neural networks. *Scientific Reports*, *13*, 13377.

Makhija, K. S., Maloney, S., & Norton, R. (2015). The utility of serial blood film testing for the diagnosis of malaria. *Pathology*, *47*(1), 68–70.

Manescu, P., Bendkowski, C., Claveau, R., Elmi, M., Brown, B. J., Pawar, V., Shaw, M. J., & Fernandez-Reyes, D. (2020a). A weakly supervised deep learning approach for detecting malaria and sickle cells in blood films. In *International conference on medical image computing and computer assisted intervention* (pp. 226–235).

Manescu, P., Shaw, M. J., Elmi, M., Neary-Zajiczek, L., Claveau, R., Pawar, V., Kokkinos, I., Oyinloye, G., Bendkowski, C., Oladejo, O. A., F, O. B., Clark, T., D, T., Shawe-Taylor, J., Srinivasan, M. A., Lagunju, I., Sodeinde, O., Brown, B. J., & Fernandez-Reyes, D. (2020b). Expert-level automated malaria diagnosis on routine blood films with deep neural networks. *American Journal of Hematology*, *95*(8), 883–891.

Mehanian, C., Jaiswal, M., Delahunt, C., & Thompson, C. (2017a). Computerautomated malaria diagnosis and quantitation using convolutional neural networks. In *Proceeding IEEE international conference on computer vision workshops* (pp. 116–125).

Mehanian, C., Jaiswal, M., Delahunt, C., Thompson, C., Horning, M., Hu, L., McGuire, S., Ostbye, T., Mehanian, M., Wilson, B., Champlin, C., Long, E., Proux, S., Gamboa, D., Chiodini, P., Carter, J., Dhorda, M., Isaboke, D., Ogutu, B., Oyibo, W., Villasis, E., Tun, K. M., Bachman, C., & Bell, D. (2017b). Computer-automated malaria diagnosis and quantitation using convolutional neural networks. In *International conference on computer vision workshops* (pp. 116–125).

Mehrtash, A., Wells, W. M., Tempany, C. M., Abolmaesumi, P., & Kapur, T. (2020). Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE Transactions on Medical Imaging*, *39*(12), 3868–3878.

Molle, V. P., Verbelen, T., Boom, C. D., Vankeirsbilck, B., Vylder, J. D., Diricx, B., Kimpe, T., Simoens, P., & Dhoedt, B. (2019). Quantifying uncertainty of deep neural networks in skin lesion classification. In *Uncertainty for safe utilization of machine learning in medical imaging and clinical image-based procedures* (pp. 52–61).

Mujahid, M., Rustam, F., Shafique, R., Montero, E. C., Alvarado, E. S., Diez, I. T., & Ashraf, I. (2024). Efficient deep learning-based approach for malaria detection using red blood cell smears. *Scientific Reports*, *14*, 13249.

Narnhofer, D., Effland, A., Kobler, E., Hammernik, K., Knoll, F., & Pock, T. (2022). Bayesian uncertainty estimation of learned variational MRI reconstruction. *IEEE Transactions on Medical Imaging*, *41*(2), 279–291.

Organization, W. H. (2020). Basic malaria microscopy – Part I: Learner's guide. Second edition. Technical Report World Health Organization.

Organization, W. H. (2022). World malaria report 2022. Technical Report World Health Organization.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, Cybernetics*, *9*(1), 62–66.

Park, H. S., Rinehart, M. T., Walzer, K. A., Chi, J. T. A., & Wax, A. (2016). Automated detection of p. falciparum using machine learning algorithms with quantitative phase images of unstained cells. *PloS ONE*, *11*(9), 1–19.

Prince, E. W., Ghosh, D., Gorg, C., & Hankinson, T. C. (2023). Uncertainty-aware deep learning classification of adamantinomatous craniopharyngioma from preoperative mri. *Diagnostics*, *13*(6), 1132.

Purnama, I. K. E., Rahmanti, F. Z., & Purnomo, M. H. (2013). Malaria parasite identification on thick blood film using genetic programming. In *International conference on instrumentation, communications, information technology and biomedical engineering* (pp. 194–198).

Qin, X., Wang, Z., Bai, Y., Xie, X., & Jia, H. (2020). Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 11908–11915). (*vol. 34*).

Quinn, J. A., Andama, A., Munabi, I., & Kiwanuka, F. N. (2014). Automated blood smear analysis for mobile malaria diagnosis. Mobile Point-of-Care Monitors and Diagnostic Device Design.

Quinn, J. A., Nakasi, R., Mugagga, P. K. B., Byanyima, P., Lubega, W., & Andama, A. (2016). Deep convolutional neural networks for microscopy-based point of care diagnostics. In *Proceedings of the 1st machine learning for healthcare conference* (pp. 271–281).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *IEEE conference on computer vision and pattern recognition* (pp. 779–788).

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster r-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149.

Rosado, L., Correia da Costa, J. M., Elias, D., & Cardoso, J. S. (2016a). Automated detection of malaria parasites on thick blood smears via mobile devices. *Procedia Computer Science*, *90*, 138–144.

Rosado, L., Costa, J. M. C. D., Elias, D., & Cardoso, J. S. (2016b). Automated detection of malaria parasites on thick blood smears via mobile devices. *Procedia Computer Science*, *90*, 138–144.

Sedai, S., Antony, B., Mahapatra, D., & Garnavi, R. (2018). Joint segmentation and uncertainty visualization of retinal layers in optical coherence tomography images using bayesian deep learning. In *Computational pathology and ophthalmic medical image analysis* (pp. 219–227).

Sensoy, M., Kaplan, L., & Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. In *International conference on neural information processing systems* (pp. 3183–3193).

Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., & Zhang, Y. (2021). TransMIL: Transformer based correlated multiple instance learning for whole slide image classification. In *Advances in neural information processing systems* (pp. 2136–2147). (*vol. 34*).

Shen, Y., & Cremers, D. (2022). Deep combinatorial aggregation. In *Advances in neural information processing systems* (pp. 32299–32310). (*vol. 35*).

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International conference on learning representations*.

Sukumarran, D., Hasikin, K., Khairuddin, A. S. M., Ngui, R., Sulaiman, W. Y. W., Vythilingam, I., & Divis, P. C. S. (2024a). An optimised YOLOv4 deep learning model for efficient malarial cell detection in thin blood smear images. *Parasites and Vectors*, *17*(1), 188.

Sukumarran, D., Loh, E. S., Khairuddin, A. S. M., Ngui, R., Sulaiman, W. Y. W., Vythilingam, I., Divis, P. C. S., & Hasikin, K. (2024b). Automated identification of malaria-infected cells and classification of human malaria parasites using a two-stage deep learning technique. *IEEE Access*, *12*, 135746–135763.

Tang, P., Yang, P., Nie, D., Wu, X., Zhou, J., & Wang, Y. (2022). Unified medical image segmentation by learning from uncertainty in an end-to-end manner. *Knowledge-Based Systems*, *241*, 108215.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*. (*vol. 30*).

Vijayalakshmi, A., & Rajesh, K. B. (2020). Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications*, *79*, 15297–15317.

Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Ieee conference on computer vision and pattern recognition* (pp. 7794–7803).

Warhurst, D. C., & Williams, J. E. (1996). Laboratory diagnosis of malaria. *Journal of Clinical Pathology*, *49*, 533–538.

Yang, F., Poostchi, M., Yu, H., Zhou, Z., Silamut, K., Yu, J., Maude, R. J., Jaeger, S., & Antani, S. (2020). Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE Journal of Biomedical and Health Informatics*, *24*(5), 1427–1438.

Yunda, L. (2011). Automated image analysis method for p-vivax malaria parasite detection in thick film blood images. *Sistemas y Telemática*, *10*(20), 9–25.

Zhang, Z., Zhang, H., Zhao, L., Chen, T., Arık, S. O., & Pfister, T. (2022). Nested hierarchical transformer: Towards accurate, data-efficient and interpretable visual understanding. In *AAAI conference on artificial intelligence*.

Zhao, O. S., Kolluri, N., Anand, A., Chu, N., Bhavaraju, R., Ojha, A., Tiku, S., Nguyen, D., Chen, R., Morales, A., Valliappan, D., Patel, J. P., & Nguyen, K. (2020). Convolutional neural networks to automate the screening of malaria in low-resource countries. *Peer Journal*, *8*, e9674.