IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 25, 2020, accepted April 14, 2020, date of publication April 27, 2020, date of current version May 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2990497

Counting and Classification of Malarial Parasite From Giemsa-Stained Thin Film Images

WASU KUDISTHALERT[1], KITSUCHART PASUPA[1], (Senior Member, IEEE), AND SISSADES TONGSIMA[2]

[1]Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand
[2]National Biobank of Thailand, National Science and Technology Development Agency, Pathum Thani 12120, Thailand

Corresponding author: Kitsuchart Pasupa (kitsuchart@it.kmitl.ac.th)

This work was supported by the Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang.

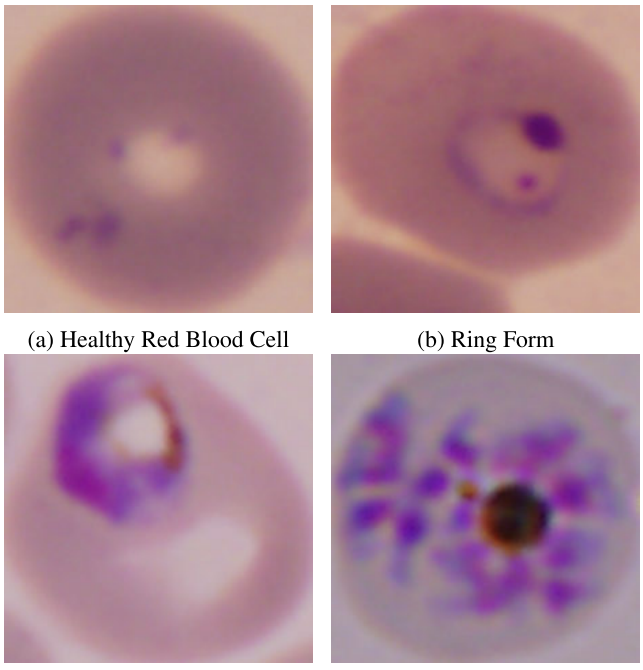
ABSTRACT Malaria is a life-threatening disease causing by an infection of the protozoan parasite *Plasmodium*. *Plasmodium falciparum* is the deadliest and most common human infected parasites hosted by anopheles mosquito vector. To cure a malaria infected patient and prevent further spreading, malaria diagnosis using microscopy to visualize Giemsa-stained parasites is commonly done. The microscopy diagnosis is somewhat time consuming and requires well-trained malaria experts to interpret what they see under the microscope. To address this limitation, an automated malaria infected diagnosis is needed. This work proposed a computer-aided automated diagnosis system that can perform remote field diagnosis with high accuracy while requiring less computational demands. The proposed framework consists of two main parts that are red blood cell counting and parasite life-cycle stage classification. The counting process is performed by computer vision techniques, namely Hough transform. Different machine learning techniques, i.e., Multilayer Perceptron, Linear Discriminant Analysis, Support Vector Machine, and Weighted Similarity Extreme Learning Machine, are employed in the classification task. We also demonstrated that combining hand-crafted and deep-learned features can enhance the overall performance of the framework. The experimental results showed that the proposed methods could correctly count and classify at 97.94% and 98.12% accuracy, respectively. The overall proposal system can achieve at 96.18% accuracy. This is achieved by WELM in conjunction with deep-learned (AlexNet_FC7) and the hand-crafted (color) features.

INDEX TERMS Combining features, Giemsa-stained thin film, malaria.

I. INTRODUCTION

Global management of malaria is very difficult since early detection of malaria infection relies mostly on diagnosis using manual microscopy of Giemsa-stained infected cells. The disease is hard to detect, especially in rural areas and developing countries due to lacking of equipment and human resources. It is a cause of millions of deaths every year [1]. The disease is caused by parasites of the genus *Plasmodium* [2]. There are four species of *Plasmodium*: *P.falciparum*, *P. vivax*, *P. ovale*, and *P. malariae*. The direness of the disease is that its symptoms are similar to an ordinary cold, e.g., fever, headache, and vomiting. In this research, we aimed to focus on *P.falciparum* that is the most commonly found and dangerous disease in Thailand. *P.falciparum* causes anemia, coma, and progresses to death if correct treatment is not considered or the diagnosis comes in very late [3]. The disease can be transmitted by female Anopheles mosquito. Infection occurs after a person was bitten by one of them. The malaria parasites—merozoites—expeditiously invade the host red blood cells (RBCs) then proceed to more developed forms. There are three stages of merozoites development that are Ring Form, Trophozoite, and Schizont (Figure 1). In the last stage, infected RBCs are filled with many Schizonts and burst open. Afterward, some merozoites will find some new RBCs and start a new life cycle. Some merozoites may develop into gametocytes after they had enough nutrients from the host. Then gametocytes can infect another host by the transmission of a vector to continue surviving as a species. There is no effective vaccine for treating malaria due to the high diversity and adaptability of *Plasmodium* antigens that make a specific vaccine developmental process not very practical [4];

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.



(a) Healthy Red Blood Cell (b) Ring Form

(c) Trophozoite (d) Schizont

FIGURE 1. Examples of stained objects.

therefore, it is imperative to detect malaria in an infected person early and recognize the degree of its spread throughout the body as well as to quantify the number of parasites and their stages in order to formulate an effective treatment [5].

Currently, the gold standard for malaria diagnosis is manual microscopy examination [6]. The diagnostic material and equipment are just some simple and multipurpose tools in a laboratory [7], therefore, the diagnosis cost is low. Although there is an availability of new methods, they require special tools that are impractical to use in some places and more expensive than the conventional technique [8]–[10]. There are two different methods for detecting and enumerating parasites: chemically staining the parasites in a thin or a thick blood smear on a slide of blood droplets and observing the visible, stained *Plasmodium* parasites. A thick blood smear method allows us to identify the morphology of the parasites easily due to the large volume examination and the concentration of dehemoglobinized RBCs. Thus, the parasites become visibly distinctive. However, it is unable to count the RBCs that are invaded by parasites. Some trophozoites and gametocytes loss during the staining process because RBCs are destroyed. This method is suitable for discovery a low parasitaemia blood slide. On the other hand, a thin blood smear method preserves the shape of RBCs and is used to visualize a parasite within a cell. Thus it is easy to count RBCs that are infected. [11], [12]. It takes around 15 minutes for an expert to manually evaluate and count 100 cells in a thin blood smear specimen [13], and the accuracy of the diagnosis directly depends on the experience of the expert [12]. In other words, this method may be unreliable when it is performed by a non-expert—the evaluation results are particularly prone to human error [14], e.g., a 100% accuracy achieved by an expert may dramatically drop down to 10% [15].

There are three main tasks in getting a diagnosis of malaria: first, an examination of the presence/absence of malaria parasites in a blood sample; second, an examination of the species of the detected parasites; and third, identification of the stages of the malaria parasites. Presently, the traditional technique is to use light microscopy to find the density and stages of malaria parasites in the blood of a patient by a microscopist, which is very time-consuming and requires great expertise. However, this task can be done by using computer vision and machine learning techniques, *in silico* processes for analyzing images such as segmentation, recognition or identification of parasites in images, without any need for an expert microscopist. There are some researches showed that computer vision techniques can be applied to segment, separate overlapping, and count RBCs in blood specimen images [16], [17]. After RBCs were identified, machine learning techniques were used to distinguish the type of RBCs, for example, healthy RBCs or infected RBCs. Moreover, they can identify the stage of infected RBCs as well.

II. RELATED WORKS

We proposed a novel computer-aided diagnosis system for malaria infection classification. This proposed system operates by processing images of Giemsa-stained thin blood film specimens and automatically reports a status of malaria infection. The advantage of this system over human experts is that it offers more consistent diagnosis under the same set of criteria, hence giving more reliable results than human experts who may judge a film specimen according to their subjective opinion. The system is based on computer vision and machine learning techniques for segmentation and classification tasks. We summarized the techniques that have been applied for this domain shown in Table 1.

There are many techniques for RBC segmentation from background for blood film images. Gatc *et al.* [21] applied thresholding technique with two thresholds, first, to segment RBCs from their background and, second, to segment parasites if RBCs were infected. Savkare *et al.* [19] and Ma *et al.* [20] applied k-means and k-medians clustering to cluster foreground and background from grayscale images by setting number of groups equal to two which are foreground and background. That technique performed well especially with blurred images. However, there are a lot of computational iterative calls for clustering pixels in an image. Ruberto *et al.* [16] applied morphological approach with two morphological operators: one that was hemispherical disk-shaped to enhance the roundness and compactness of RBCs and the other one that was flat disk-shaped to separate overlapping cells. These two hemispherical and flat disk-shaped structuring elements work together with some knowledge of the structure of RBCs to be able to segment individual and also overlapping RBCs in images. Sharif *et al.* [22] attempted to apply a watershed transform

TABLE 1. Summary computer-aided techniques for malaria diagnosis.

Segmentation	Feature Extraction	Classifying Method
Circular Hough transform [17] [18]	Bag of visual words [23]	Convolutional Neural Networks [27] [28]
k-means and k-medians clustering [19] [20]	Color channel intensity [24]	K-nearest neighbor [29]
Morphological approach [16]	Color, Shape and Texture histogram [25] [26]	Multilayer perceptron [26]
Thresholding approach [21]	Raw image [27] [28]	Naive Bayes [30]
Watershed transform [22]		Support Vector Machine [27]
		Trees Classifier [28]

on a distance map together with a concavity analysis in order to determine the split lines for overlapping cells. That method could only be applied to clumped cells of only two RBCs. For clumped cells that contain more than two cells, they used a template matching technique to find individual RBCs. This technique can separate small cells but demanded high computational cost. Maitra *et al.* [17] and Mahmood and Mansor [18] used a circular Hough transform in an automatic application for segmentation and to automatically count the number of RBCs in an image. That technique can identify and separate overlapping RBCs within only one process. Nevertheless, that technique requires users to specify the radius of the circle of RBC to be detected, but this radius can be calculated from the size of the foreground object that is expected to be a RBC in an image. All in all, to find the right radius is much easier than to find a good brightness intensity threshold, for thresholding and watershed transform, so we chose to use Hough transform for the mentioned task in this study.

After RBCs were detected, feature extraction technique was used to encode RBC images into features. This technique plays a major role in image classification. These features were expected to contain the relevant information which is the key to conduct an effective classification model. López-Puigdollers *et al.* [23] reported that constructing an expert system utilizing local image descriptor techniques that could recognize six white blood cell types. They built a bag of visual words with a keypoint detector and regular sampling techniques. They found that using the Oriented Features from an Accelerated Segmentation Test as an interest point detector to localize keypoints on cell contours and regular sampling strategies yielded a better classification performance. However, there are simple and practical methods that are often used, for example, Markiewicz *et al.* [25] and Ross *et al.* [26] used a basic user-defined common image features (hand-crafted features) such as a histogram of color, shape or texture, but Habibzadeh *et al.* [27] and Quinn *et al.* [28] used only raw images as features for deep Convolutional Neural Networks (CNN). It should be noted Yeon *et al.* [24] suggested to employ the green component as a representative color space for blood film image because it gave a higher contrast and a greater range of color intensities than red and blue components which is easier to distinguish objects in an image and the performance of feature extraction depends on quality of images as well. Recently, Pasupa *et al.* [31] showed that using hand-crated features together with deep-learned features, a features extracted from CNN, as multiple representations of data gave a better performance than using individual features.

Machine learning technique used to distinguish classes by learning from sample data which is supervised learning. There are many machine learning techniques that were used for classifying malaria parasite and its stage [29]. Ross *et al.* [26] and Díaz *et al.* [30] demonstrated some useful applications of standard machine learning methods in malaria image diagnosis, i.e., k-nearest neighbor classifier (kNN), Naïve bayes, Support Vector Machine (SVM) and Multilayer perceptron (MLP) together with histogram of hand-crafted features such as texture, shape and color to classify malarial RBCs. Díaz *et al.* [30] also revealed that kNN and MLP along with normalized RGB color space gave better performance than the others. In 2012, the beginning of deep learning era, Krizhevsky *et al.* [32] designed a CNN architecture, known as AlexNet which is the well-known CNN architecture, to compete in the ImageNet Large Scale Visual Recognition Challenge and won several international competitions. CNN is a deeper version of multilayer perceptron which consists of several convolution and fully connected layers. It is commonly applied to analyze image data because it can automatically extract features and classify image within itself. Nowadays, CNN has been applied to classify images from bio-medical domains, for example, Quinn *et al.* [28] used CNN to recognized Microscopy-Based Point of Care Diagnosis such as diagnosis of malaria in thick blood smears, tuberculosis in sputum samples, and intestinal parasite eggs in stool samples comparing with tree classifier and evaluated with area under the receiver operating characteristic curve and average precision. Their results showed that CNN was a robust method. It yielded higher performance in all cases due to well trained image model constructed by a large training set. Habibzadeh *et al.* [27] tried to apply CNN to classify RBCs. The CNN took an entire image as input. Its classification results were compared to those of several standard methods including SVM. It was found that CNN yielded a higher rate of recognition than SVM-based classifiers did. As it can be seen, those studies used raw images and hand-crafted features such as shape, size, color and texture as pertinent data for classifying healthy/infected blood cell classes or to differentiate parasite developmental stages. Those basic data were satisfactory for classifying those classes and stages. However, deep-learned features extracted from deep-learned models were recently

used instead of those basic features [28]. In addition, hand-crafted and deep-learned features can be complementary to each other [31]. Thus, combining these features might improve the overall performance for diagnosis of human malarial infection. In addition, there are reports that show that current machine learning techniques are able to deal with highly imbalanced datasets [33]–[35] and the lack of labeling problem [36] in biomedical domain.

In this paper, we propose a machine learning technique to deal with several malaria RBC classes, i.e., Healthy/Infected RBCs and the malarial parasites at different growth stages by learning from images of samples of blood cell types and malarial infection stages. We used Weighted Similarity Extreme Learning Machine (WELM). It is an improved version of one of the powerful classification methods named Extreme Learning Machine (ELM) [37] which based on single-hidden layer feedforward neural networks. The weights of the model can be calculated without iterative process due to weights of the model obtained from performing linear regression on data mapped through non-linear projection. There are some evidences that showed the potential of using WELM classification method on one of the most challenging datasets, Maximum Unbiased Validation Dataset, which is dramatic imbalance between classes and highly diverse in structure. The experiment result showed that WELM give better performance than other standard methods, i.e., SVM, Random Forest, and Similarity Searching [38]–[40]. In this study, we demonstrated a list of contribution as follow:

- We proposed a framework to count and classify Healthy/Infected RBCs from Giemsa-stained thin film images.
- The proposed framework was tested on a proposed malarial giemsa-stained thin blood film image dataset which is a highly imbalance dataset and, moreover, it contains more than 23,000 malarial RBC images.
- Deep-learned features was used as model features which was extracted from a pre-trained of AlexNet.
- Evaluate the performance of using multiple types of feature sets, i.e., hand-crafted features, deep-learned features, and also concatenated features. We show that hand-crafted feature can be complementary to deep-learned feature.
- Comparing the performance of using WELM and three comparable standard classification methods, i.e., MLP, SVM, and Linear Discriminant Analysis (LDA).
- The proposed system works like an expert doing the whole malaria diagnosis process. The whole process included 1) feeding image samples into the system, which is the same as an expert observing blood slides under a microscope; 2) detecting and counting RBCs in a blood slide; and 3) recognizing classes of RBC, for example, class of Healthy RBC, Ring Form, Trophozoite or Schizont. For the system, the accuracy of the whole process is then evaluated.

III. METHODS

In this section, we explain the methods used in this experiment. Morphological image processing and Hough transform were used as image pre-processing techniques. Morphological image processing technique was used to clean noises and reconstruct object structures. Hough transform technique was used to segment RBC objects. Transfer learning technique and Weighted Similarity Extreme Learning Machine are used to perform deep feature extraction and classify the RBCs.

A. MORPHOLOGICAL OPERATION

A morphological operation is an image processing technique based on the shapes of objects in an image. It is a pre-processing step before the pre-processed image is processed further and used for noise removal and restoration of original shapes of objects from distorted ones. A morphological operation applies a structuring element s to an object. The structuring element is a template of small shape, e.g., a square, disk or rectangle which is applied to a binarized input image (of which each pixel has a value of either 0 or 1). There are two main morphological operators: erosion $\ominus$ and dilation $\oplus$ [41]. For erosion, the boundary pixels of every foreground object that are adjacent to the background is converted to the background. We can define erosion that produces a new binary image g from an input binary image f as:

$$g = f \ominus s. \tag{1}$$

For example, $g(x, y) = 1$ at a structuring element's origin coordinate when the structuring element fits exactly to the input object in an image, and it is 0 otherwise. Conversely, for dilation, the boundary pixels of every foreground object that are adjacent to the background is converted to the foreground when the structuring element touch or intersect the object. Dilation is expressed as:

$$g = f \oplus s. \tag{2}$$

Figures 2b and 2c show how erosion and dilation operators can eliminate some noise and fill the hole in an object, respectively. In addition, examples of opening and closing operations adapted from the two common operators, erosion and dilation [42], [43]. The equation that represents the closing operator performs dilation followed by erosion indicated in Figure 2d. It can be expressed as

$$g = (f \oplus) \ominus s. \tag{3}$$

On the other hand, the opening operator performs erosion then dilation indicted in Figure 2e. It can be expressed as

$$g = (f \ominus) \oplus s. \tag{4}$$

B. HOUGH TRANSFORM

Hough transform is a conventional feature extraction technique for the detection of straight lines in images [44]. It has been extended to detect some low-parametric forms such as circles and ellipses [45], [46]. The technique finds some

(a) Input image (b) Eroded image (c) Dilated image (d) Closed image (e) Opened image

FIGURE 2. Morphological filters.

required parameters of the desired object in an image by a voting procedure on an accumulator array A, well-known as Hough space. A simple equation of a straight line is,

$$y = mx + c. \tag{5}$$

We can rearrange the equation in order to find c to be

$$c = -xm + y, \tag{6}$$

where m is a gradient or a slope of the straight line and c is the y-intercept. This equation was applied to an image by varying parameter m in the accumulator array A, in the m-axis and c-axis in Hough space, and a local maxima was obtained for each coordinate (x, y). However, when the straight line becomes a vertical line–90 degrees angle to the x-axis–the slope becomes unbound. Therefore, Hesse normal form was brought into deal with this problem:

$$\rho = x\cos\theta + y\sin\theta, \tag{7}$$

where ρ is the distance from the origin to closest point on the straight line and θ is the angle between the x-axis and the line. Hence, the accumulator consists of pairs of coordinates on the ρ and θ axes. With the same procedure, the technique could be applied to find circle in images through the following equation:

$$(x-a)^2 + (y-b)^2 = r^2, \tag{8}$$

where (a, b) is the center of the circle and r is the radius for every coordinate (x, y) that appears in the image. On the other hand, in some applications, the parameter r is known, thus the equation with a fixed parameter r can be expressed as the following system of equations:

$$a = x - r\cos\theta, \tag{9}$$

$$b = y - r\sin\theta. \tag{10}$$

The most voted candidate in the accumulator array A indicates a center of the circle that represents RBC. The voting procedure for circle detection is represented in Algorithm 1.

In this artificial intelligence era, many applications in computer vision domain employ machine learning techniques, a kind of statistical techniques, for solving some problems such as recognition and classification of objects. The purpose of the technique is to learn patterns from sample data as experience with no specific explicit instructions in the program to do so. This technique utilizes mathematical optimization to

Algorithm 1 Circular Hough Transform

```
function HOUGH_TRANSFORM(x, y, r)
    for t ← 1 to 360 do          ▷ vary all possible angles θ
        a ← x-r * cos(t * π/180)
        b ← y-r * sin(t * π/180)
        A[a, b, r] ← A[a, b, r] + 1          ▷ voting
    end for
end function
```

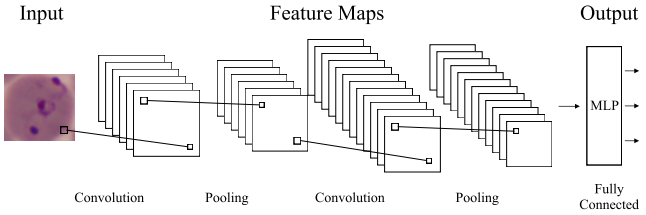


FIGURE 3. An example of the architecture of a convolutional neural network.

do supervised learning tasks such as classification or prediction and unsupervised learning tasks such as clustering from the test data [47]–[49]. In this work, we used a deep neural network and computer vision techniques to extract features to represent images. Then the extracted features will be fed to the model to classify types of RBCs as well as infection stages of malarial parasites.

C. TRANSFER LEARNING

The origin of deep convolutional neural network architecture–the neocognitron–was proposed by [50]. The idea was proposed a long time ago, but it was not popular due to the lack of advanced high-performance computing facility. Currently, Graphics Processing Units (GPUs) have rapidly evolved to become high-performance computing tools. Therefore, deep learning techniques have become popular due to their promising performance. The most common type of this technique is CNN. The architecture of CNN, as shown in Figure 3, is a combination of the following layers:

- Convolution layer extracts features from an input image as a feature map by computing the dot product between the input image and a receptive field that slides across the width and height of the entire image.
- Pooling layer performs a down-sampling operation that is a non-linear function. A pooling window slides on the image with non-overlapping steps to reduce the size

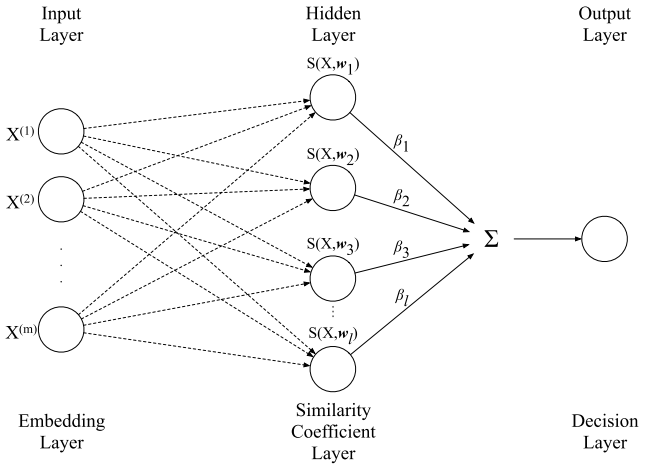


FIGURE 4. Weighted Similarity Extreme Learning Machine architecture.

of the feature map and yield higher-level features. For example, a conventional pooling layer, a max pooling function, chooses the maximum value in each small region as a representative value.

- The fully connected layer consists of neurons that fully connect to all activations in the previous layer–similar to those in the conventional MLP. The fully connected layers need to classify the high-level features obtained from the convolution layers and pooling layers into classes.

However, it still requires a high computational cost to build a CNN model from scratch. To ease the problem, transfer learning is utilized. Transfer learning is the idea to apply the knowledge that is learned from the previous task to tackle a new related task when there is a limited number of training samples. Using this concept results in a more reliable model and faster training speed than training a model from scratch [51]. A common way is to use a pre-trained model—was trained on a larger dataset—on a similar task. There are many popular pre-trained models for recognition tasks, e.g., AlexNet [32], VGG [52], and YOLO [53].

Instead of training our own CNN model, we use a pre-trained model to extract features from malarial RBCs images in this work. To do that, we conserve the convolutional layers for feature extraction task and replace the last layer—MLP—with Weighted Similarity Extreme Learning Machine instead.

D. WEIGHTED SIMILARITY EXTREME LEARNING MACHINE

Weighted Similarity Extreme Learning Machine (WELM) is based on Extreme Learning Machine (ELM) that was first proposed by [37]. ELM was modified from a generalized single-hidden layer feedforward neural network (SLFN) in which solves the bottleneck problem—neural-by-neural weight calculation—of SLFN by randomly selection weights from the training samples and process them through a generalized inverse function. Thus, hidden layer of ELM can be trained without an iterative process. WELM architecture consists of m input neurons which are the number of input dimensions. The input neurons are fully connected with l hidden neurons which are weights of the model $\boldsymbol{w}_i$. The set of weights connecting the hidden layer to the output layer is denoted as $\boldsymbol{\beta}$. Therefore, the model can be defined as:

$$y_j = \sum_{i=1}^{l} \beta_i h(\boldsymbol{x}_j), \tag{11}$$

where $\mathrm{H} = [h(\boldsymbol{x}_1), \ldots, h(\boldsymbol{x}_n)]$ is the hidden layer output for an input X where $\mathrm{X} \in \mathcal{R}^{n \times m}$ corresponding to target labels $\boldsymbol{y}$, and n is the number of input samples. The H matrix for WELM can be expressed by a similarity activation function $s(\cdot)$, such as a Euclidean similarity function, with a linear combination of input samples X and weights W as follows:

$$\mathrm{H} = \begin{bmatrix} h(\boldsymbol{x}_1) \\ \vdots \\ h(\boldsymbol{x}_n) \end{bmatrix} = \begin{bmatrix} s(\boldsymbol{x}_1, \boldsymbol{w}_1) & \ldots & s(\boldsymbol{x}_1, \boldsymbol{w}_l) \\ \vdots & \ddots & \vdots \\ s(\boldsymbol{x}_n, \boldsymbol{w}_1) & \ldots & s(\boldsymbol{x}_n, \boldsymbol{w}_l) \end{bmatrix}_{n \times l}, \tag{12}$$

where W is randomly selected from X instead of generated by a continuous probability distribution. This is to ensure that the weights are in same the distribution and dimension span. Therefore, W is a subset of X and $l \leqslant n$. The prediction score can be expressed by following equation:

$$\hat{\boldsymbol{y}} = \mathrm{H}\boldsymbol{\beta}, \tag{13}$$

WELM needs to minimize the approximation error expressed as the square root error below,

$$\min_{\boldsymbol{\beta}} \frac{1}{2} \parallel \mathrm{H}\boldsymbol{\beta} - \boldsymbol{y} \parallel_2^2 . \tag{14}$$

Thus, the optimal solution of $\boldsymbol{\beta}$ can be expressed by Moore-Penrose pseudo-inverse function as follows,

$$\boldsymbol{\beta} = (\mathrm{H}^{\mathrm{T}}\mathrm{H})^{-1}\mathrm{H}^{\mathrm{T}}\boldsymbol{y}. \tag{15}$$

The architecture of WELM is shown in Figure 4 and the procedure of WELM for malarial classification is shown as pseudo-code in Algorithm 2.

IV. PROPOSED SYSTEM

The proposed system is shown in Figure 5. It consists of eight stages; the first four were stages of image pre-processing processes; the next two were stages of RBCs segmentation processes; and the last two involved classification task. Each stage is explained in more details below.

1) Image quality enrichment: input images were first transformed into binary images (shown in Figure 6b) which were then put through a morphological filter, i.e., a disk-shaped opening filter, which performs erosion then dilation, removed some noise, e.g., the artefacts shown in Figure 6c; then a flood-fill operation filled holes in the RBCs as shown in Figure 6d; the holes in the cells are a usual distinctive characteristic of RBCs due to the lack of hemoglobin at the small central area of the cells [54].

Algorithm 2 Weighted Similarity Extreme Learning Machine

function WELM_TRAIN(X, $\boldsymbol{y}$)
  W ← randomly select subset of X
  n ← #samples in X
  n_H ← #samples in Healthy Red Blood Cell class
  n_R ← #samples in Ring Form class
  n_T ← #samples in Trophozoite class
  n_S ← #samples in Schizont class
  for $i \leftarrow 1$ to n **do**
    $\gamma_i = \sqrt{\frac{\max(n_H, n_R, n_T, n_S)}{\#\text{samples in } y_i \text{ class}}}$
  end for
  $\hat{\mathrm{H}} \leftarrow \boldsymbol{\gamma} \cdot \mathrm{S}(\mathrm{X}, \mathrm{W})$
  $\boldsymbol{\beta} \leftarrow (\frac{1}{C} + \hat{\mathrm{H}}^{\mathrm{T}}\hat{\mathrm{H}})^{-1}(\hat{\mathrm{H}}^{\mathrm{T}}\boldsymbol{\gamma} \cdot \boldsymbol{y})$
  return W, $\boldsymbol{\beta}$
end function
function WELM_PREDICT(W, $\boldsymbol{\beta}$, $\mathrm{X}_{\mathrm{Test}}$)
  $\mathrm{H} \leftarrow \mathrm{S}(\mathrm{X}_{\mathrm{Test}}, \mathrm{W})$
  $\hat{\boldsymbol{y}} \leftarrow \mathrm{H}\boldsymbol{\beta}$
  return $\hat{\boldsymbol{y}}$
end function

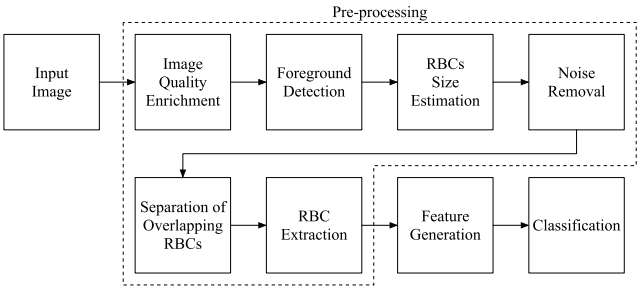


FIGURE 5. The workflow of the proposed approach.

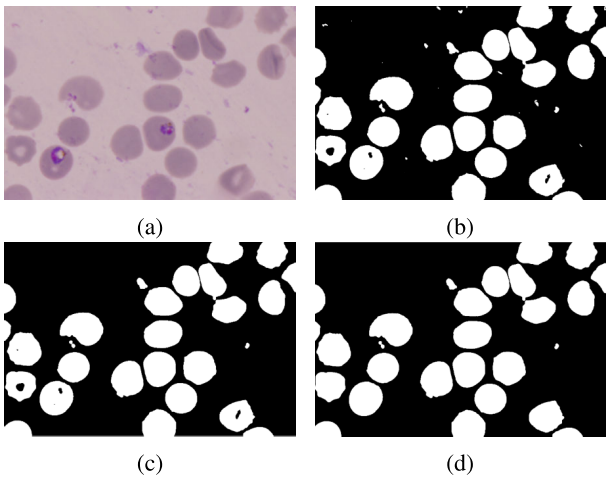


FIGURE 6. The workflow stages of the proposed approach–image quality enrichment: (a) input image; (b) transforming the input image into a binary image; (c) removing noise from the binary image; (d) filling the holes inside the RBCs.

2) Foreground detection: a regional maxima technique was used to identify objects in the image. This technique took an object to be represented by an 8-connected component, thus the white pixels in the binary image that were adjacent to the edges or corners of the object were taken to be a part of the object [55]. At the end of this step, the technique collected all foreground objects, i.e., individual and overlapping RBCs and some unwanted objects, e.g., noise and artefacts that still had not been removed; hence some objects were still ambiguous. Figure 7 shows the foreground and background of the processed image. We then got rid of all unwanted objects in the next stage.

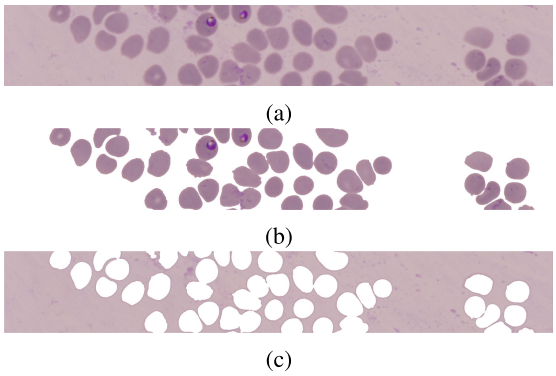


FIGURE 7. A foreground and background detection by the regional maxima: (a) original image, (b) foreground, and (c) background.

3) RBCs size estimation: The proposed approach was capable of differentiating the types of objects presented in an image: artefacts as well as non-overlapping and overlapping RBCs. In order for the method to be able to differentiate them, it needed to know a common size for an RBC. However, sizes of RBCs can be different; hence we needed to estimate the size range of these RBCs that were suitable for our purpose. This could be done by using a frequency histogram, which plotted RBC sizes, to differentiate the sizes of objects. Since some RBCs would be present in an image taken for our purpose, the RBC size that exhibited the maximum frequency in the histogram was selected as the representative size, for example, as shown in Figure 8, 22,245.75 pixels is used as a RBC size for only a considering blood field image. We attempted to find an optimum range of RBC size by trial and error and found that the range of ±35% of the representative size was optimum. Objects not within this range were taken to be abnormal objects. It should be noted that the frequency is the number of occurrences of RBC objects in an image and the blood cell size is the representative number of pixels of RBC objects in the image.
4) Noise removal: the objects that were smaller than the smallest size in the defined range were considered as noise, and they were removed immediately. On the other hand, the objects that were bigger than the biggest size in the defined range were collected to analyze further whether they were artefacts or overlapping RBCs in the next stage.

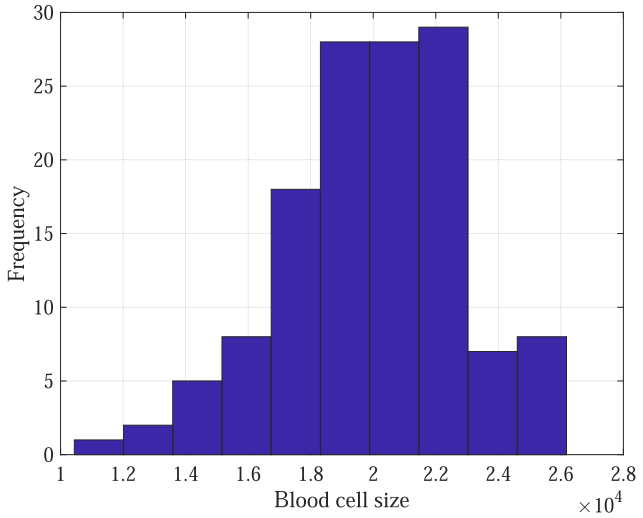


FIGURE 8. Histogram of RBC sizes from a field image.

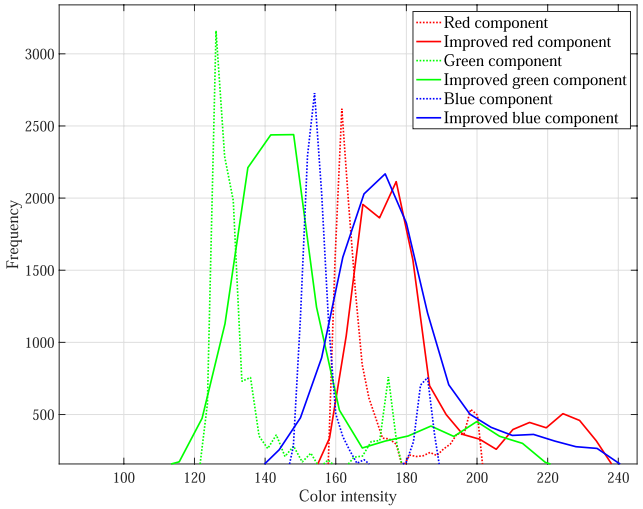


FIGURE 9. Color intensity histogram of each channel in RGB components; dotted lines represent the original color intensities while solid lines represent the improved color intensities by CLAHE.

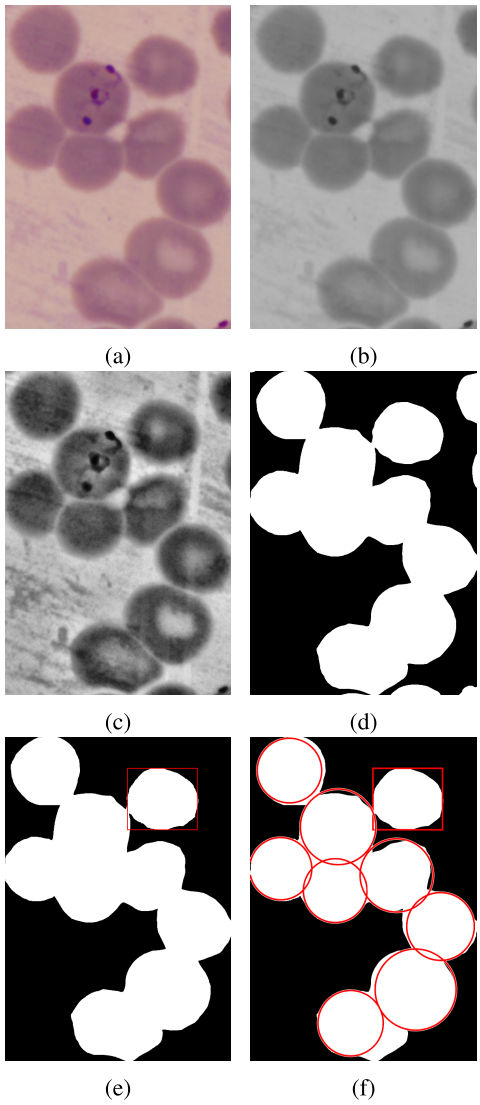


FIGURE 10. Separation process of overlapping RBCs: (a) abnormal region; (b) transforming the region into a greyscale image; (c) improving the contrast of the image by CLAHE; (d) binarizing the image; (e) iteratively applying an opening filter; (f) detecting circular objects in the image.

5) Separation of overlapping RBCs: Contrast-limited adaptive histogram equalization (CLAHE) was employed to enhance the contrast of each object [56]. This technique could sharpen the edges of RBCs so that the boundary of an individual RBC could be distinguished clearly. As shown in Figure 9, the green component gave the highest contrast and a greater variety of color intensities. On the other hand, the red and blue components gave a lower contrast and a smaller range of color intensities; therefore, the green component was chosen as a representative color channel [24]. The difference in contrast can be seen in Figure 10a, 10b and 10c. Next, segmentation of overlapping RBCs involved two steps. First, we applied an opening filter iteratively until the object was separated into two or more new objects. Those separated objects then were determined whether their sizes were in the defined range or not. If they were so, they were collected as non-overlapping RBCs, else they were removed. If the overlapping RBCs still could not be separated, the second step was applied: a Hough transform was used to identify circles in the region as shown in Figure 10f. Since the Hough transform technique had a parameter that needed to be specified, i.e., the radius of the RBC (r) is simply calculated by the square root of the RBC area (size) divided by π. In this work, we set r between $\pm 35\%$ of the optimal RBC size to collect all of possible blood cells in images.
6) RBCs extraction: all RBC objects were cleaned, separated and identified by the previous stage, and in this stage, the regional maxima technique was used once more to collect and count the individual RBCs.

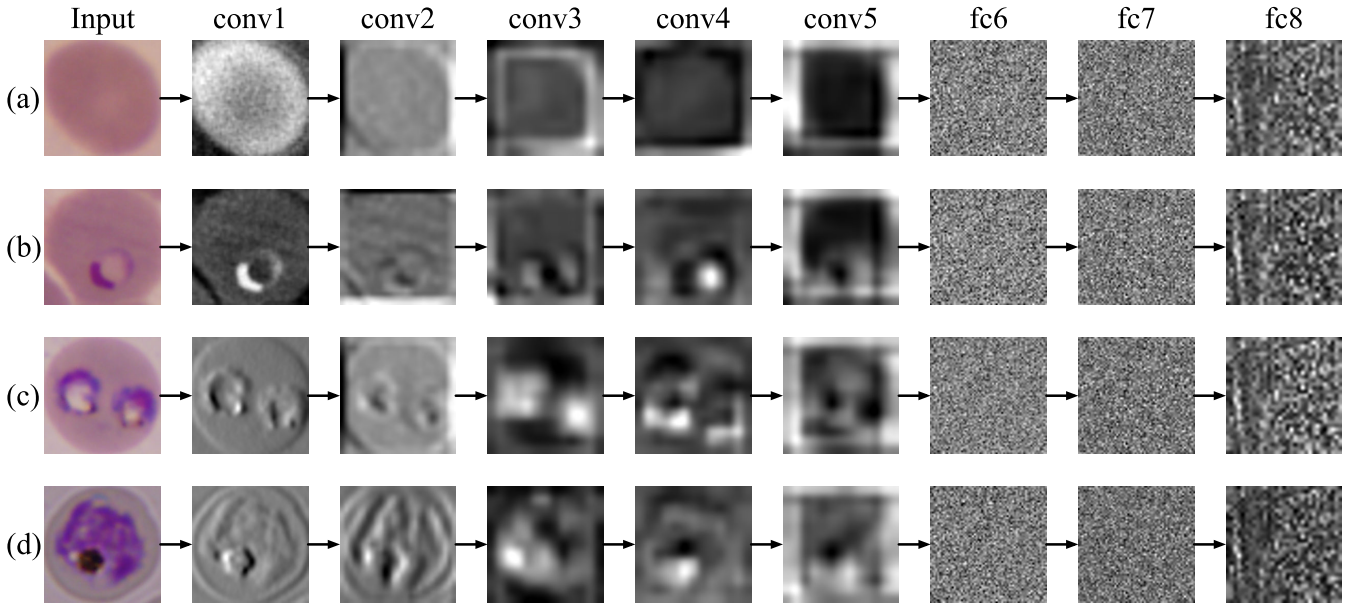


FIGURE 11. Feature extraction throughout the AlexNet network: (a) Healthy RBC; (b) Ring Form; (c) Trophozoite; (d) Schizont.

7) Feature generation: Raw images of RBCs were collected. The images then needed to be differentiated into classes but they contained only color information that could not be used directly and efficiently as features. Hence, this stage applied methods that enabled interpretation of raw images into useful features. In this experiment, we had four candidates for useful features: color, shape, texture and deep-learned feature. Because each type of feature gave different kinds of information, it was necessary to find which feature provide more information for the classification of RBCs. For example, the color feature might be a truly useful one since the color of Healthy RBCs image and that of RBCs in the Schizont class were visually different. For the deep feature, we extracted this feature by a pre-trained convolutional neural network called AlexNet [32] of which architecture is often used as the base architecture for other convolutional neural networks. AlexNet was still a great choice. Among other competitive CNN models, it yielded very good results [57]. The original pre-trained AlexNet is eight layers deep and has been trained on over millions of images from the ImageNet database [58]. The examples of features obtained from the AlexNet along the network are shown in Figure 11. In this work, we used two deep features from the 7^{th} and 8^{th} fully connected layer that contained 4,096 and 1,000 features, respectively, denoted as FC7 and FC8 (the architecture is shown in Figure 12) because the 7^{th} layer provided a high number of features but the 8^{th} layer provided higher-level features. It should be noted that a higher layer of Deep CNN provide more abstract information and semantics features. In contrast, a lower layer provide information that are similar to image details [59]. This work also used three 1024-bin histograms of image features: color, shape and texture. The shape feature was extracted with a Sobel technique which is generally used for identifying vertical and horizontal gradients in images. A Gabor technique with filter directions of 0, 45 and 90 degrees was used for collecting texture feature. For the feature map in Figure 13, we reshaped a vector of feature into a square image to illustrate the pattern of data in each class generated by a different feature generator technique. For example, the color histogram technique provided a pattern of color frequency. The color of the RBC in the Healthy RBC and Ring Form classes yielded a certain pattern that represents the distinctively red color while the more purplish color of the cell infected with a malarial parasite at a more mature stage in the Trophozoite and Schizont classes are represented by a differently looking pattern of color frequency.

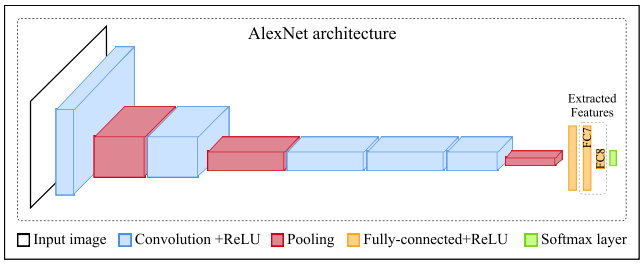


FIGURE 12. The AlexNet architecture that provided two deep features from fully-connected layers denoted as FC7 and FC8.

8) Classification: the last stage of the workflow is classification of RBCs into the right classes: Healthy RBCs, Ring Form, Trophozoite and Schizont. In this

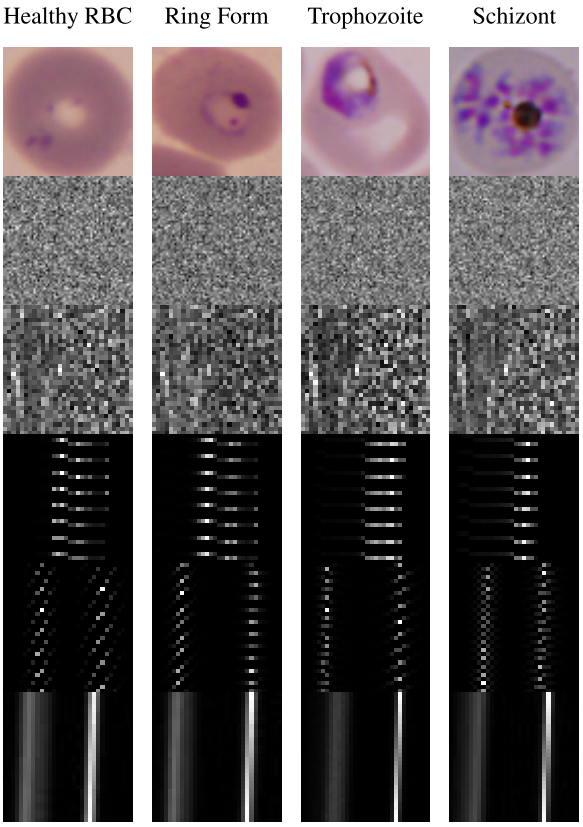


FIGURE 13. Feature map where the columns are malarial stages and the rows are features ordered as follow: input image; AlexNet_FC7; AlexNet_FC8; color histogram; shape histogram; texture histogram.

paper, we proposed two classification frameworks to handle this task denoted by M1 and M2, as shown in Figure 14. The M1 approach, a one stage classification, classifies RBCs in one stage similar to the way malarial classification expert does. Thus, only one classification model was needed to be trained as shown in Figure 14a. On the other hand, the M2 approach is a two-stage tree classifier [26] two separate classification models were required to perform two sequential tasks: a model for classifying infected/uninfected RBCs in first stage and another model for classifying malarial life-stages in the second stage, as shown in Figure 14b. Thus, the first stage of the M2 approach dealt only with a binary task which was easier to deal with than a multi-class problem. The second stage of the M2 approach classified only three classes while the M1 approach classified four classes.

V. EXPERIMENTAL FRAMEWORK

A. IMAGE ACQUISITION

Images of *P.falciparum* parasitized RBC were taken at the Protein-Ligand Engineering and Molecular Biology laboratory at the National Center for Genetic Engineering and Biotechnology (BIOTEC), Pathum Thani, Thailand. Thin blood smears for malaria had been prepared for drug susceptibility testing of anti malarial agents. Well trained researchers

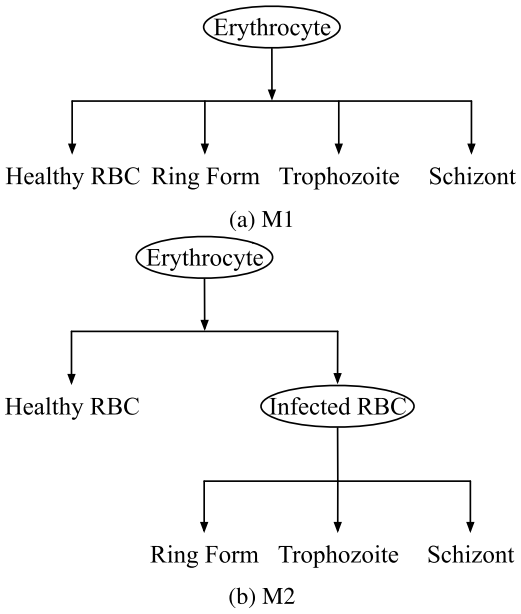


FIGURE 14. Classification approaches.

TABLE 2. Detail of training set and test set.

Dataset	Malarial infected stages				Total
	Healthy RBC	Ring Form	Trophozoite	Schizont	
Training set	20,061	393	294	175	20,923
Test set	2,229	44	33	19	2,325
Total	22,290	437	327	194	23,248

then counted, by using thin blood smear to accentuate the parasites, the numbers of healthy RBCs and infected RBCs. The images of samples used in this experiment were only of erythrocytes, malarial parasites and some artefacts from the removal of white blood cells from in vitro cultured samples. Examples of these images are in Figure 15.

The images were acquired using Olympus BX51 microscope and a DP71 digital camera system. The specimens were examined under oil immersion with a 100× physical magnification and a 10× optical zoom. The images were captured at a resolution of 4080 × 3072 pixels in TIFF image file format. We were able to collect 147 field-of-view images. After RBC segmentation, we obtained a dataset with the size of 23,248 RBC images. The dataset consisted of 22,290 Healthy RBCs, 437 Ring Forms, 327 Trophozoites and 194 Schizonts as shown in Table 2. It can be seen that the dataset is highly imbalance one with a ratio of 1:24 for infected and Healthy RBCs as shown in Figure 17. It is noted that 0.7% of the total RBCs accounting for ambiguous RBCs caused by poor slide preparation were removed from the dataset. Poor slide preparation can cause artefact, damage, and/or poor morphology to the cells [60]. Especially, the T-distributed Stochastic Neighbor Embedding (t-SNE) technique [61] was used to visualize high-dimensional data into a low-dimensional space of two dimensions. The dataset is then visualized the distribution in Figure 16. The dataset is available upon request.

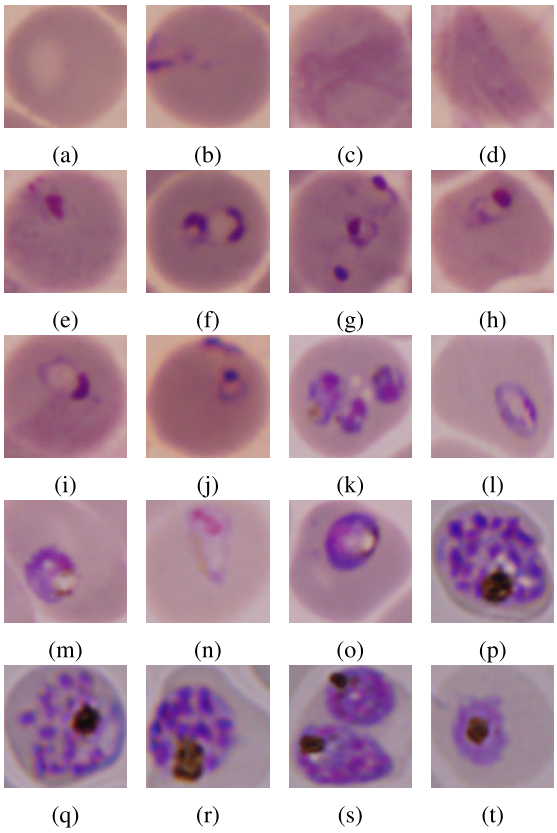


FIGURE 15. Stained objects in the dataset: (a) Healthy RBC, (b)-(e) artefacts, (f)-(j) Ring Form, (k)-(o) Trophozoite and (p)-(t) Schizont.

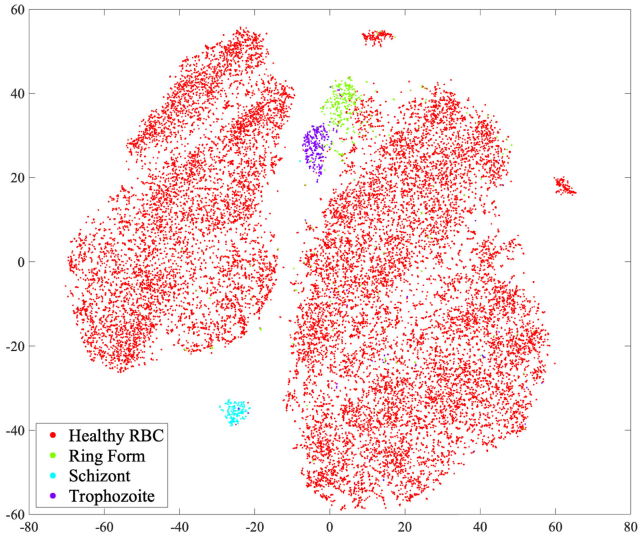
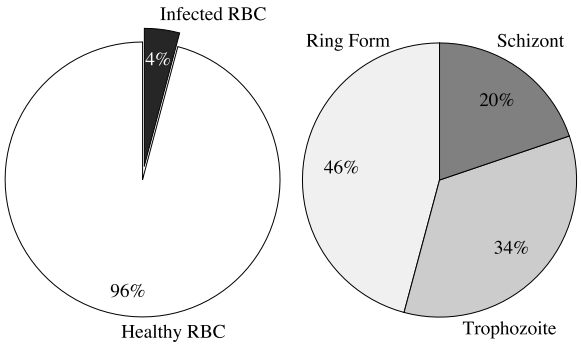


FIGURE 16. Data distribution of the dataset.

B. PROPOSED FRAMEWORK

In this study, we divided the evaluation of the proposed method into three parts:

- The performances of the classification methods were evaluated. We demonstrated that using multiple feature representations, especially using a combination of user-defined image features and deep-learned features, would give a better performance than using either one of the described feature sets. We discussed the results of feature evaluation by WELM and obtained the top three features that yielded the highest evaluation metric scores. Furthermore, these top three features were used and evaluated by four selected methods, which are MLP, LDA, SVM, and WELM.
- The performances of the whole process (both counting and classification stages) was evaluated. The best method and its features obtained from the previous part were used.
- Furthermore, the most robust method was compared with various conventional methods on an additional Maralia dataset [57]. The dataset consists of 27,558 samples of Parasitized and uninfected cells.



(a) Healthy vs. Infected RBCs (b) Three infected stages

FIGURE 17. Proportions of classes in dataset.

C. EXPERIMENTAL SETTINGS

In this experiment, we divided the dataset into training and test sets. The size of the training set was 90% of the entire dataset and the size of the test set was 10%. A simple random sampling approach, which is commonly used, efficient, and easy to implement [35], [62], [63], was used to partition the data into ten sets with different random seeds. We reported the average scores and standard deviation values of these ten runs.

The data in the training set comprised 20,923 samples: 20,061 samples of Healthy RBCs and 862 samples of infected RBCs with identical proportion of members in each class on whole dataset. Noting that, the set was highly imbalance, i.e., 96% of all data in the training set were data of the Healthy RBCs class, which is the easiest class to classify. Therefore, we needed to reduce the burden of computational cost and fixed the imbalance problem by randomly undersampling samples from the class of Healthy RBCs such that the number of samples in this class was cut down to 862 samples, the same number as the number of samples in the infected RBCs class. Hence, the total number of samples in the

adjusted training set was only 1,724. It is noted that the samples in the three infected classes remained the same and were used entirely to train the model due to the limited number of samples for those infected classes in the training set. Thus, the class ratio of the training set was 862:393:294:175 for Healthy:Ring Form:Trophozoite:Schizont that still caused an imbalance problem.

In this study, we utilized an efficient classification method, WELM to classify malarial blood-stained images into four classes, namely, Healthy RBC, Ring Form, Trophozoite and Schizont. WELM is extremely fast, easy to implement, able to work with highly imbalanced dataset, highly accurate, and especially consumes low computational resources. Thus, this method can be practically used in the field where computational resources are limited. In this paper, we used MATLAB on MacBook Pro (Model: Late 2013), having a 2.4 GHz CPU, 256 gigabytes of storage and 8 gigabytes of RAM for implementing experiments.

We evaluated WELM's performance against three standard classification methods: MLP, LDA, and SVM. Each of these methods has different weaknesses and strengths. For example, MLP is one of the most widely used classification methods. Nowadays, MLP have been embedded in many deep convolutional neural network architectures. LDA is the simplest method that deals with linear data, so it can be trained rapidly and there is no hyperparameter to be tuned. SVM is currently the most widely used classifying method for numerous applications. It utilizes support vectors that act as a dividing hyperplane in feature space for classifying data into classes. Lastly, WELM is a classifying method based on a single layer feed forward neural network that can be trained much faster than any common ANN. It also uses a kernel trick like the SVM, i.e., using a similarity kernel to transform an existing feature space into a new feature space. In addition, it uses a pseudoinverse technique to determine the weights for the model that provide the lowest error rate.

As for parameter settings, there are three methods of which parameters need to be tuned. For MLP algorithm, two parameters need to be tuned: the number of hidden layers and the number of hidden nodes. For our setting, we set the number of hidden layers to be one because the image dataset was small but the number of features was high, too many layers might lead to a model overfitting problem and waste higher computational resources [64] that may not be available in a field work. The number of hidden nodes was set to [8, 16, 32, 64, 128, 256, 512]. For SVM, only one parameter was tuned: regularizing C. For this experiment, it was set to $[10^{-6}, 10^{-5}, \ldots, 10^{5}, 10^{6}]$. For WELM, two parameters were tuned: regularizing C which was set to be $[10^{-5}, 10^{-4}, \ldots, 10^{4}, 10^{5}]$ and the percentage of hidden nodes which was in the range of $[10, 20, \ldots, 90, 100\%]$. We then used a ten-fold cross validation procedure to find the optimal parameters for each experiment. Thus, all parameters would be changed depending on the data that we obtained from random picking for each experiment. For the kernel tricks used in these methods, a linear kernel was embedded in LDA and SVM while a Euclidean kernel was embedded in WELM. Since WELM works with similarity weight, for a fair comparison, we selected Euclidean kernel, which is a linear kernel that works with similarity weight. Both of these kernels are based on an algebraic operation called dot product [65].

TABLE 3. Abbreviations of features.

Abbreviation	Feature
DF1	AlexNet_C7
DF2	AlexNet_C8
IF1	Color
IF2	Shape
IF3	Texture

TABLE 4. Abbreviations of concatenated features.

Abbreviation	Feature
CF1	DF1, IF1
CF2	DF1, IF2
CF3	DF1, IF3
CF4	DF2, IF1
CF5	DF2, IF2
CF6	DF2, IF3
CF7	IF1, IF2
CF8	IF1, IF3
CF9	IF2, IF3
CF10	DF1, IF1, IF2
CF11	DF1, IF1, IF3
CF12	DF1, IF2, IF3
CF13	DF2, IF1, IF2
CF14	DF2, IF1, IF3
CF15	DF2, IF2, IF3
CF16	IF1, IF2, IF3
CF17	DF1, IF1, IF2, IF3
CF18	DF2, IF1, IF2, IF3

VI. RESULTS AND DISCUSSIONS

A. EVALUATION THE PERFORMANCE OF CLASSIFICATION METHODS

We illustrated the performance of comparable methods, features and classification approaches, particularly, the most robust method in this section. The experimental results are the average of 10 runs and of two approaches, M1 and M2. Tables 3 and 4 list the abbreviations of features used in both experiments. These features are divided into three groups: Deep features (DF), Image features (IF) and Concatenated features (CF).

1) A COMPARISON OF FEATURES PERFORMANCE

Table 5 lists the performance metrics achieved by the average across two classification approaches, M1 and M2, with 23 different features sets. The WELM algorithm was evaluated with various kinds of metrics in order to show the degrees of different capabilities of the model clearly. These metrics are as follows.

- Accuracy: accuracy is a standard metric which is the percentage of the number of correct predictions; however, it is not good for evaluating imbalance datasets because the numbers of samples in different classes in this kind

TABLE 5. Performance metrics achieved by each set of features. There are three types of features: deep features, image features and concatenated features. The metric for the best feature of each type is underlined, while the metric for the best feature among all types of features is marked in bold.

	Accuracy	AUC	MCC	Sensitivity	Specificity	F_1-score	Rank-sum
DF1	**98.15 ± 0.38**	99.50 ± 0.30	83.75 ± 2.05	94.99 ± 1.39	98.91 ± 0.41	87.36 ± 2.11	212
DF2	97.92 ± 0.34	**99.61 ± 0.34**	82.15 ± 2.38	93.75 ± 1.83	98.66 ± 0.52	86.12 ± 2.08	193
IF1	87.25 ± 1.20	94.91 ± 1.25	54.95 ± 2.23	86.58 ± 2.46	93.19 ± 1.01	65.74 ± 2.01	38
IF2	90.20 ± 0.61	87.56 ± 2.15	35.97 ± 1.88	55.71 ± 2.59	93.03 ± 0.72	47.58 ± 1.99	10
IF3	94.95 ± 0.47	98.96 ± 0.40	69.61 ± 2.50	91.21 ± 2.17	98.00 ± 0.42	76.07 ± 2.30	68
CF1	98.12 ± 0.25	99.49 ± 0.30	**84.09 ± 1.88**	95.51 ± 1.65	98.66 ± 0.48	**87.94 ± 1.60**	**216**
CF2	97.55 ± 0.30	98.28 ± 0.58	78.22 ± 2.30	90.09 ± 2.00	98.01 ± 0.62	82.82 ± 2.04	137
CF3	97.55 ± 0.37	99.51 ± 0.29	81.65 ± 1.72	95.50 ± 1.32	**98.92 ± 0.35**	85.92 ± 1.35	191
CF4	98.07 ± 0.31	99.56 ± 0.25	83.84 ± 2.06	95.08 ± 1.69	98.58 ± 0.50	87.77 ± 1.65	213
CF5	97.24 ± 0.36	98.62 ± 0.48	75.18 ± 2.38	88.24 ± 2.34	97.79 ± 0.57	80.17 ± 2.19	95
CF6	97.50 ± 0.38	99.57 ± 0.32	81.05 ± 2.34	95.07 ± 1.63	98.91 ± 0.35	85.35 ± 1.90	174
CF7	91.36 ± 0.87	95.39 ± 0.52	55.74 ± 2.59	84.74 ± 2.42	94.90 ± 0.54	65.11 ± 2.32	32
CF8	95.83 ± 0.33	99.05 ± 0.29	72.79 ± 2.14	94.34 ± 0.99	98.38 ± 0.38	78.55 ± 2.12	78
CF9	91.40 ± 0.56	96.49 ± 1.01	49.09 ± 3.57	73.53 ± 4.95	97.15 ± 0.45	57.30 ± 3.27	21
CF10	97.64 ± 0.28	98.70 ± 0.44	79.51 ± 2.04	92.88 ± 2.00	98.12 ± 0.52	84.10 ± 1.77	161
CF11	97.34 ± 0.37	99.57 ± 0.23	80.67 ± 2.02	**95.99 ± 1.13**	98.91 ± 0.38	85.15 ± 1.68	172
CF12	96.54 ± 0.37	99.29 ± 0.27	74.80 ± 2.25	92.11 ± 2.37	98.43 ± 0.43	80.12 ± 1.99	98
CF13	97.46 ± 0.33	98.94 ± 0.29	77.96 ± 1.94	92.36 ± 1.70	97.98 ± 0.49	82.79 ± 1.70	129
CF14	97.29 ± 0.37	99.60 ± 0.22	80.45 ± 2.02	95.93 ± 0.88	98.91 ± 0.28	84.98 ± 1.74	170
CF15	96.46 ± 0.46	99.14 ± 0.30	73.42 ± 3.63	90.36 ± 3.10	98.42 ± 0.46	78.76 ± 3.24	82
CF16	94.75 ± 0.28	98.61 ± 0.41	63.84 ± 2.11	90.59 ± 1.57	98.10 ± 0.37	69.60 ± 2.02	49
CF17	96.84 ± 0.38	99.29 ± 0.27	76.31 ± 2.32	93.98 ± 1.56	98.72 ± 0.27	81.23 ± 2.05	109
CF18	96.87 ± 0.37	99.32 ± 0.23	76.46 ± 2.25	93.68 ± 1.77	98.69 ± 0.34	81.39 ± 1.93	112

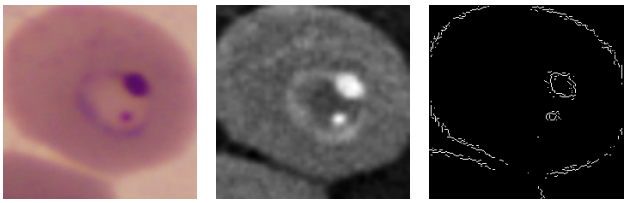
of dataset may be widely different, thus it is possible that the accuracy value may be influenced strongly by only one large class in the dataset.

- AUC: the area under the receiver operating characteristic curve (AUROC) is the area under a curve of true positive rate versus false positive rate that can be used well as a metric for methods operating on imbalance datasets.
- MCC: Matthews correlation coefficient (MCC) is a metric that was formulated for evaluating the quality of binary classification; it was designed to be used as a balanced measure that can be used on imbalance datasets. The equation takes into account all true positive (TP), true negative (TN), false positive (FP) and false negative (FN) results. MCC can be directly calculated from a confusion matrix as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \tag{16}$$

- Sensitivity: some call this metric a TP rate or recall; it is the rate of correctly identified TP predictions to all positive outcomes.
- Specificity: some call this metric a TN rate or selectivity; it is the rate of correctly identified negative predictions to all negative outcomes.
- F_1-score: this score is a harmonic mean of precision and sensitivity; it can be used as an overall performance metric.
- Rank-sum: it is the sum of the ranks that a tested feature could achieve in several runs; twenty three candidates were ranked in an ascending order in 10 runs (observations), which the lower performance results was ranked first to the higher performance results, thus the ideal rank-sum is 230. This rank-sum order would be validated by a statistical significance method called 'Kruskal-Wallis one-way analysis of variance by ranks test H' or 'one-way ANOVA on ranks'. This validating method was used for testing whether candidates were from identical populations or not by comparing the medians of the groups [66]. In this experiment, we used F_1-scores for overall score measurement; they were the input into the Kruskal-Wallis method. The output of the method was a statistically significant level of $p < 0.001$. Thus, each feature in this experiment had a different continuous distribution significantly.

As shown in Table 5, using each DF1 and DF2 individual feature as a deep feature provided nearly the same outcomes for every kind of metrics. This was not so with image features where the IF3 texture feature provided a better outcome compared to those provided by the individual color or shape



(a) Color image (b) Deep-learned image (c) Shape image

FIGURE 18. Processed images of a malaria-infected RBC.

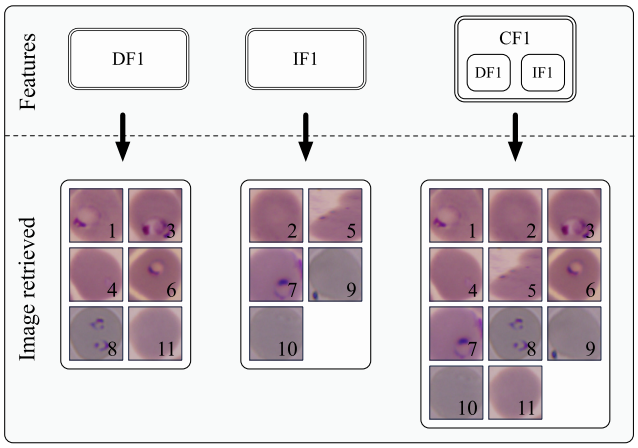


FIGURE 19. Comparative performances of individual and concatenated features.

feature. On the other hand, the CF1, CF4 and DF1 features, respectively, gave the best outcomes, CF1 combining a deep feature with color features—DF1 and IF1—got the highest scores for almost all of the metrics. Note that using the deep feature together with the color feature as input tended to give a better classification outcome than using the other features or combinations of features because the deep feature provided information that was similar to that provided by shape features, deep features can provide more information than shape features can, for example, RBC contour and shapes of malarial parasites and also shape of surface of objects as shown in Figure 18b. Thus, using it together with the color feature, the available information was more complete and more useful. Figure 19 illustrates an example from the experimental results showing that using a concatenated feature consisting of the two features mentioned above provided a better classification outcome than any single individual feature could. Specifically, by using the concatenated feature consisting of DF1 and IF1, the entire set of images were retrieved, but only some of the images were retrieved when any individual feature alone was used. Another point to be noted is that the shape feature, IF2, tended not to provide a good result. The shape feature was able to capture the shape of RBCs but only some parts of malarial parasites were detected which had the same circular shape with RBC but smaller, which might be ascribed the underlying deficiency to the shapes of all blood cells being not very different, and so the outcome was vague as shown in Figure 18c.

TABLE 6. Abbreviations of combinations of M1 and M2 approaches of four methods and the top three features.

Method		Feature		
		DF1	CF1	CF4
MLP	M1	MLP_{M1}-DF1	MLP_{M1}-CF1	MLP_{M1}-CF4
	M2	MLP_{M2}-F1	MLP_{M2}-F1	MLP_{M2}-F4
LDA	M1	LDA_{M1}-DF1	LDA_{M1}-CF1	LDA_{M1}-CF4
	M2	LDA_{M2}-DF1	LDA_{M2}-CF1	LDA_{M2}-CF4
SVM	M1	SVM_{M1}-DF1	SVM_{M1}-CF1	SVM_{M1}-CF4
	M2	SVM_{M2}-DF1	SVM_{M2}-CF1	SVM_{M2}-CF4
WELM	M1	$WELM_{M1}$-DF1	$WELM_{M1}$-CF1	$WELM_{M1}$-CF4
	M2	$WELM_{M2}$-DF1	$WELM_{M2}$-CF1	$WELM_{M2}$-CF4

2) A COMPARISON OF WELM WITH THE BEST THREE FEATURE SETS AGAINST OTHER METHODS

The previous section reports the features that gave the best classification outcomes. This section reports the results of using the top three feature sets with some of the standard classification methods: MLP, LDA, SVM, and WELM. In order to make the results listed in every table and figure hereafter easier to follow, a list of abbreviated indexes identifying every method is shown in Table 6.

Figure 20 shows an error plot of F_1-scores achieved by all methods listed in Table 6. Please note that the scores and standard deviations achieved by methods under M1 and M2 approaches are shown as blue and red lines, respectively and that the three best features described in section VI-A1–CF1, CF4 and DF1–that gave the best results in Table 5 were used in the experiment in Section VI-A2. It can be observed that the $WELM_{M1}$-CF1 model achieved the highest F_1-score at 87.95 ± 1.79. In contrast, the MLP_{M1}-CF4 model achieved the lowest score at 66.67 ± 11.13. The most stable model was LDA_{M1}-DF1 which achieved an F_1-score of 86.01 and a very low standard deviation of ±1.08. It should also be noted that the models can be clearly separated into two groups: a group of MLP models with a mean F_1-score in the range of 66–73 and another group of models from other methods with a mean score in the range of 85–88. A possible reason that MLP models did not perform as well as models from other methods was that MLP typically needs to learn from a large amount of data for each malarial stage and this study investigated four malarial stages but there was not a lot of data for each stage. Another reason was that the size of the training features was at least 1,000 variables for DF2 and the maximum size was 7,168 variables for CF18. Therefore, the training data was high dimensional in features but its sample size was small. There was a cause of a small-sample-size problem that may lead to a high chance of overfitting due to the large size of features that did not properly match with the size of the training data [67]. Evidently, the trend that the test errors were higher than the training errors can be clearly seen in Figure 21, signifying that the models were actually overfitted and did not fully converged. All of these reasons

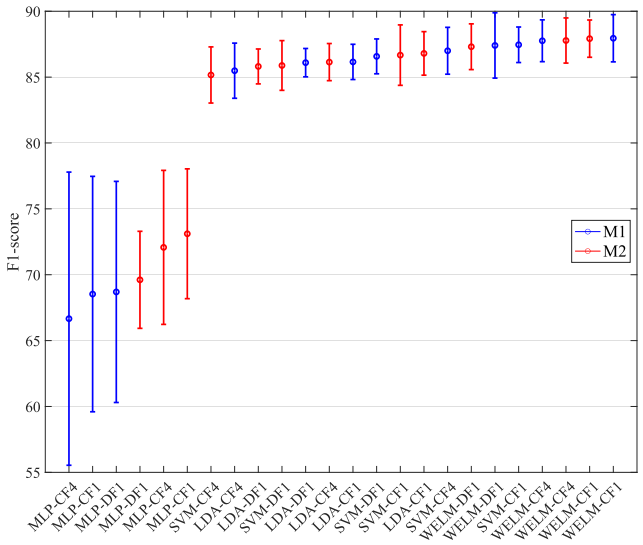


FIGURE 20. Plotting of F_1-scores together with the standard deviations of ten experiments on MLP, LDA, SVM and WELM and two approaches; the scores and standard deviations for M1 approach are blue lines while those for M2 approach are red lines.

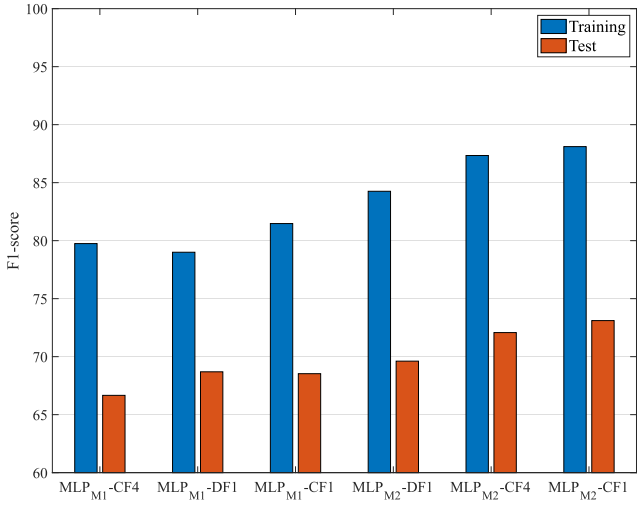


FIGURE 21. Bar charts of training and test errors of M1 and M2 approaches with MLP method.

could explain why the models from MLP did not perform well.

Figure 22 displays confusion matrices of the top and bottom two performers, $WELM_{M1}$-CF1, $WELM_{M2}$-CF1, MLP_{M1}-CF3 and MLP_{M1}-CF1, respectively. It can be seen that all four models were nearly equally successful at classification samples of the Healthy RBC class, which were the easiest to be classified, at 96–98% rate, and their retrieval accuracies for samples of the Healthy RBC class were also very close. However, the effectiveness of the models cannot be judged by the overall accuracies shown in the matrices in this table due to the imbalance of the dataset where the overall accuracies were based mostly on the 97–98% accuracy of the Healthy RBC class alone. More weight should be given to the accuracies identification Infected class identification

TABLE 7. F_1-scores of M1 and M2 approaches by four machine learning techniques on the top three features.

Method	Approach	
	M1	M2
MLP-CF1	68.53	73.11
MLP-CF4	66.66	72.08
MLP-DF1	68.69	69.61
LDA-CF1	86.15	86.80
LDA-CF4	85.49	86.14
LDA-DF1	86.10	85.81
SVM-CF1	87.46	86.67
SVM-CF4	87.00	85.16
SVM-DF1	86.57	85.88
WELM-CF1	87.95	87.92
WELM-CF4	87.76	87.78
WELM-DF1	87.40	87.31
Average	82.15	82.86

results: Ring Form, Trophozoite and Schizont. When infected classes are considered, it can be observed that every model was able to retrieve samples of Trophozoite and Schizont nearly equally well. The Ring Form class was the exception, of which the retrieval was clearly less effective. Most frequently, the models would misclassify samples of the Ring Form class as those of Healthy RBC class because the images of RBC in the Healthy RBC class often had "color-staining" artefacts from the step of blood smearing on a slide. Frequently, these artefacts could be mistaken as malarial parasites in their ring form stage and thus confused the models. This kind of confusion also happens to expert interpretation results. Examples of these problematic images are shown in Figure 23.

3) A COMPARISON OF CLASSIFICATION APPROACHES

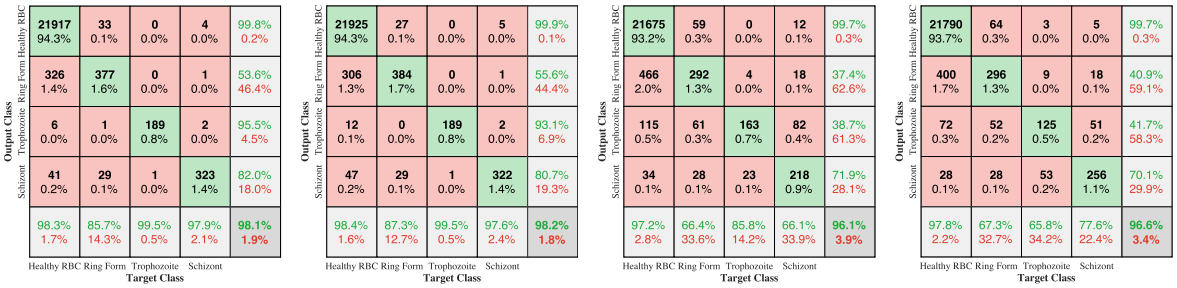
The comparative performances, in terms of F_1-score, of those M1 and M2 approaches are shown in Table 7. The 12 methods of each approach were tested 10 times. It can be seen that the 12 methods of the M1 approach achieved a higher F_1-score than the 12 methods of the M2 approach for 61 out of 120 runs. However, the mean F_1-score achieved by the methods in the M2 approach was higher than that achieved by the methods in the M1 approach. In addition, the result from a two-sample t-test indicated that the mean scores from both approaches were not significantly different at $p = 0.5161$; therefore, we concluded that M1 and M2 approaches performed equally well in terms of F_1-score.

4) A COMPARISON OF MOST ROBUST METHODS

The performances of all methods were sorted according to the sum of their F_1-score ranks of 10 runs and are shown in Figure 24. The reliability of the ranked order was statistically evaluated with a tool called Kendall's Coefficient of Concordance W,

$$W = \frac{12\sum_{i=1}^{N}\bar{R}_i^2 - 3N(N+1)^2}{N(N^2-1)}, \tag{17}$$

where $\bar{R}$ is the average ranked order of the i-th candidate and N is the number of candidate methods



(a) WELM_{M1}-CF1 (b) WELM_{M2}-CF1 (c) MLP_{M1}-CF4 (d) MLP_{M1}-CF1

FIGURE 22. Confusion matrices of the top two performers, WELM_{M1}-CF1 and WELM_{M2}-CF1, and the bottom two performers, MLP_{M1}-CF4 and MLP_{M1}-CF1.

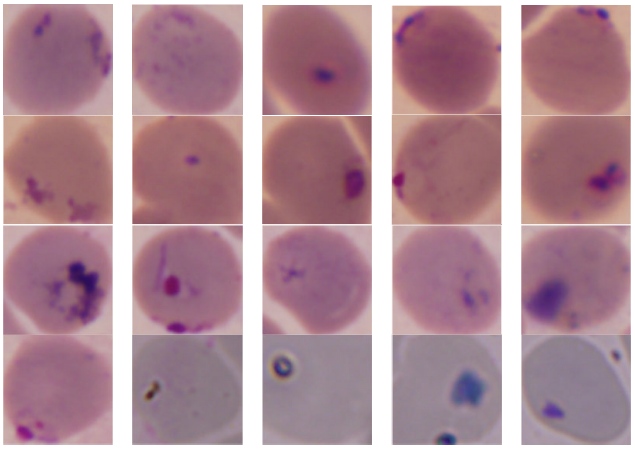


FIGURE 23. Images of Healthy RBCs with artefacts which were predicted as parasites of the Ring Form stage.

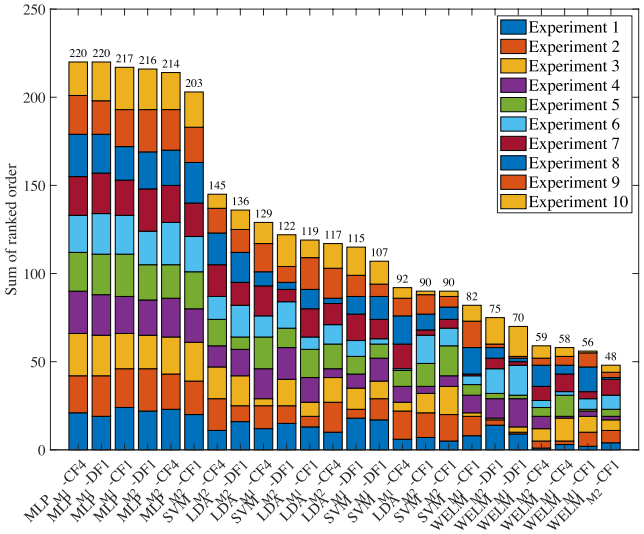


FIGURE 24. Sums of ranked order in terms of F_1-score performance, reported as stacked bars in descending order of 10 experiments.

(equals to 24 here). The value of W found was 0.6984 which was transformed into a χ value by the following equation,

$$\chi^2 = k(N-1)W, \tag{18}$$

where k is the number of experiments (equals to 10). The χ^2 value obtained was 160.64 which indicates that the ranked order shown in Figure 24 was reliable at a confidence level of 99.9%.

Furthermore, it can be seen that the most robust model in this study was WELM_{M2}-CF1 of which the sum of its ranked order was only 48 from 10 experimental runs of 24 methods, followed by WELM_{M1}-CF1, WELM_{M1}-CF4, and WELM_{M2}-CF4, respectively. The sums for these three latter methods were nearly the same at about 50. Not surprisingly, the worst methods were MLPs of which the sums were over 200 because of overfitting problem. However, the approaches (M1 and M2) of each method provided nearly the same ranked order in total agreement with the obtained classification statistics shown in Table 7.

B. EVALUATION THE PERFORMANCE OF WHOLE PROCESS OF PROPOSED SYSTEM

In this section, we evaluate the proposed system described in section IV in order to count and classify RBCs in blood film images. The best classification method obtained from Section VI-A4 was used in the proposed system.

1) EVALUATION OF RBCs COUNTING TASK

We evaluate the performance of the system on 60 randomly selected film images. The performance is reported in Table 8. It can be seen that our system can efficiently perform segmentation and count with the correct identification of 9,218 out of 9,420 RBCs that is equivalent to 97.86% accuracy. We found that the main problem of this task is the blood film staining quality. The proposed system can highlight RBCs and parasites in blood films but also misclassified those artifacts and spot of noises as well. The three percentages of error of this experiment mostly came from mistreated artifacts, as shown in Figure 23, and miscounting a clump of RBCs as one.

TABLE 8. Accuracy of counting performance of the proposed system.

Image #	System Count	Manual Count	Accuracy (%)
1	155	163	95.09
2	159	162	98.15
3	168	170	98.82
4	171	174	98.28
5	164	168	97.62
6	138	139	99.28
7	141	141	100
8	116	116	100
9	135	136	99.26
10	133	133	100
11	97	97	100
12	183	188	97.34
13	152	154	98.70
14	171	175	97.71
15	177	179	98.88
16	165	170	97.06
17	157	162	96.91
18	165	173	95.38
19	165	171	96.49
20	158	163	96.93
21	133	136	97.79
22	125	125	100
23	148	148	100
24	122	126	96.83
25	120	124	96.77
26	136	143	95.10
27	120	125	96.00
28	101	103	98.06
29	136	145	93.79
30	109	113	96.46
31	105	107	98.13
32	116	119	97.48
33	120	120	100
34	163	163	100
35	117	118	99.15
36	169	170	99.41
37	154	159	96.86
38	133	137	97.08
39	124	131	94.66
40	166	167	99.40
41	117	117	100
42	121	122	99.18
43	160	164	97.56
44	271	273	99.27
45	286	295	96.95
46	263	265	99.25
47	259	271	95.57
48	202	205	98.54
49	127	128	99.22
50	163	163	100
51	163	164	99.39
52	102	103	99.03
53	142	144	98.61
54	171	175	97.71
55	168	180	93.33
56	221	227	97.36
57	193	202	95.54
58	149	151	98.68
59	147	150	98.00
60	104	106	98.11
Average	153.63	156.95	97.94

2) EVALUATION OF RBCs CLASSIFICATION TASK

To complete the classification of malarial infected stages, the 9,218 countable RBCs from the prior process will be classified using the most robust method, $WELM_{M2}$-CF1. However, we were unable to assign infection stage classes to 19 RBCs due to ambiguity. Therefore, we allocated them to the unknown class. We reported the experimental results of the proposed system in two aspects—evaluated with and without unknown class—as shown in Table 9. The performance metrics used in this table is the accuracy of each class. It performed nearly 100% accuracy, especially on the classification of Trophozoite which is 99.64%. In overall picture, the system can achieve 95.99% and 96.18% accuracy for counting with and without unknown, respectively.

TABLE 9. Proposed system prediction accuracy.

Class	True	Countable	Recognizable	
			WELM (%)	AlexNet (%)
Healthy RBC	9,048	8,855	**8703.00 (96.19)**	8678.20 (95.91)
Ring Form	172	169	162.70 (94.59)	**167.70 (97.50)**
Trophozoite	111	110	**109.60 (98.74)**	109.10 (98.29)
Schizont	70	68	**67.00 (95.71)**	66.90 (95.57)
Unknown	19	16	0.00	0.00
Total with unknown	9,420	9,218	**9042.30 (95.99)**	9021.90 (95.77)
Total without unknown	9,401	9,202	**9042.30 (96.18)**	9021.90 (95.97)

In addition, we fine-tuned the AlexNet model by training the last three layers of the architecture to differentiate four classes i.e. Healthy RBC, Ring Form, Trophozoite and Schizont. The model was trained for 10 epochs with 8 batch sizes. The accuracy that it achieved was compared to that from the most robust model, WELM, as shown in Table 9. It can be seen that WELM yielded a higher average accuracy score for three out of four classes. Therefore, it was remarkable that AlexNet was able to recognize Ring Form class at three more percent higher accuracy.

C. EVALUATION OF THE ADDITIONAL DATASET

We have proved the performance of our method by evaluating it on another dataset from Lister Hill National Center for Biomedical Communications (LHNCBC), part of National Library of Medicine (NLM), US [57], against comparable traditional methods, i.e., WELM, MLP, SVM and ELM [68]. Two different types of the kernel function—Linear and Radial Basis Function (RBF)—were employed in SVM, here, we refer as Linear-SVM and RBF-SVM, respectively. This dataset contains 27,558 cell images and consists of two classes: Healthy RBC and Infected RBC. The dataset was made with a balanced number of samples between classes.

The experiment was run with five-fold split tests at the patient-level. The performance results were reported as confusion matrices in Figure 25. According to the Figure, WELM still performed very well—obtaining the highest overall accuracy. It yielded the highest recognition rate of Infected RBC class at 94.6%. Besides, MLP could recognize the Healthy RBC class with the same level of accuracy as WELM at 93.5%. The worst one was Linear-SVM which got a 92.3% overall score and also got the lowest recognition rate of Healthy RBC class at 90.2%. The overall score could be ranked as those from WELM, MLP, RBF-SVM, ELM and Linear-SVM, respectively. Thus, the most robust method on this dataset was WELM.

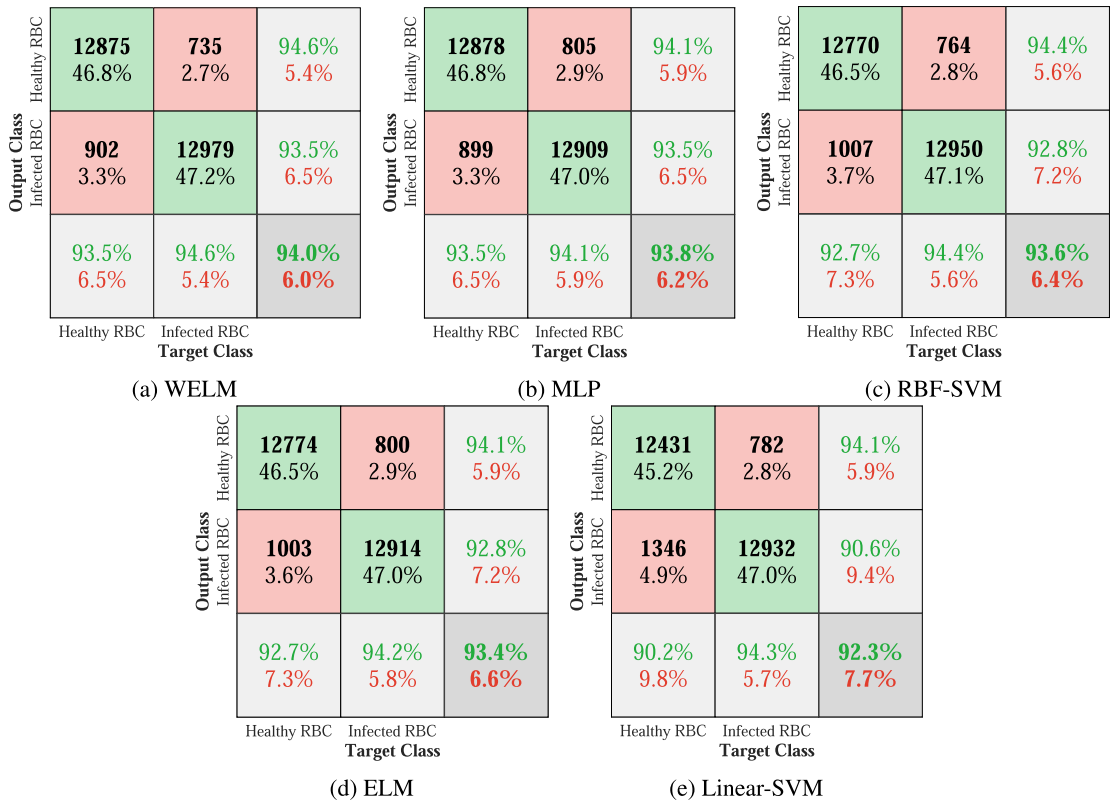


FIGURE 25. Confusion matrices of comparable methods perform on additional dataset.

VII. CONCLUSION

An automated malaria parasite detection system for a Giemsa-stained thin-film image was presented. The proposed image processing system can count and classified both healthy and malaria-infected RBCs. The system could achieve as high as 96.18% accuracy. In this work, we employed both user-defined image features (i.e., color, shape, texture) and deep-learned features in the RBCs classification task. Deep-learned features were extracted from a pre-trained deep convolutional neural network named AlexNet. The 23 different sets of features were evaluated and compared using various methods, including MLP, LDA, SVM, WELM. It was found that a combination of color feature and deep-learned (AlexNet_FC7) features could enhance the overall performance of the task, conveying the need of a combination of obvious image feature with deep-learned features. We also showed that WELM in conjunction with a combination of color and AlexNet_FC7 features is the most robust method for this task. For future work, we aim to (i) enhance our system to be able to classify more malarial species, such as *P.vivax*, *P.ovale*, and *P.malariae*, and (ii) perform semantic segmentation for RBC in order to eliminate the background instead of using standard image processing techniques that contain many tuning parameters. Furthermore, we also aim to make an explainable-AI system for malaria domain. A black-box approach and its decision cannot be explained. However, users of a system, such as doctors or patients, need to understand the origin of the problems and to retrace the decisions made by the system.

ACKNOWLEDGMENT

The authors would like to thank Ms. Parichat Prommana and Dr. Chairat Uthaipibull from the Protein-Ligand Engineering and Molecular Biology laboratory, National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand, for providing and validating Giemsa-stained thin blood films.

REFERENCES

[1] M. T. MAKLER C. J. PALMER A. L. AGER, "A review of practical techniques for the diagnosis of malaria," *Ann. Tropical Med. Parasitol.*, vol. 92, no. 4, pp. 419–433, Jun. 1998.

[2] W. E. Collins, "Plasmodium knowlesi: A malaria parasite of monkeys and humans," *Annu. Rev. Entomol.*, vol. 57, no. 1, pp. 107–121, Jan. 2012.

[3] P. K. Sarkar, G. Ahluwalia, V. K. Vijayan, and A. Talwar, "Critical care aspects of malaria," *J. Intensive Care Med.*, vol. 25, no. 2, pp. 93–103, Mar. 2010.

[4] A. F. Cowman, D. Berry, and J. Baum, "The cellular and molecular basis for malaria parasite invasion of the human red blood cell," *J. Cell Biol.*, vol. 198, no. 6, pp. 961–971, 2012.

[5] B. Nadjm and R. H. Behrens, "Malaria: An update for physicians," *Infectious Disease Clinics North Amer.*, vol. 26, no. 2, pp. 243–259, 2012.

[6] M. M. Kettelhut, "External quality assessment schemes raise standards: Evidence from the UKNEQAS parasitology subschemes," *J. Clin. Pathol.*, vol. 56, no. 12, pp. 927–932, Dec. 2003.

[7] P. B. Bloland, "Drug resistance in Malaria," World Health Org., Geneva, Switzerland, Tech. Rep. WHO/CDS/CSR/DRS/2001.4, 2001.

[8] A. Voller, D. Bidwell, G. Huldt, and E. Engvall, "A microplate method of enzyme-linked immunosorbent assay and its application to Malaria," *Bull. World Health Org.*, vol. 51, no. 2, p. 209, 1974.

[9] D. Chilton, "Use of rapid diagnostic tests for diagnosis of malaria in the UK," *J. Clin. Pathol.*, vol. 59, no. 8, pp. 862–866, 2006.

[10] G. Snounou, S. Viriyakosol, X. Ping Zhu, W. Jarra, L. Pinheiro, V. E. do Rosario, S. Thaithong, and K. N. Brown, "High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction," *Mol. Biochem. Parasitol.*, vol. 61, no. 2, pp. 315–320, Oct. 1993.

[11] M. Guy, "Basic malaria Microscopy: Part I. Learner's guide," *Trans. Roy. Soc. Tropical Med. Hygiene*, vol. 86, no. 6, p. 700, 1992.

[12] M. Pammenter, "Techniques for the diagnosis of Malaria," *South Afr. Med. J. Suid-Afrikaanse Tydskrif vir geneeskunde*, vol. 74, no. 2, pp. 55–57, 1988.

[13] Y.-K. Chan, M.-H. Tsai, D.-C. Huang, Z.-H. Zheng, and K.-D. Hung, "Leukocyte nucleus segmentation and nucleus lobe counting," *BMC Bioinf.*, vol. 11, no. 1, p. 558, 2010.

[14] K. Mitiku, G. Mengistu, and B. Gelaw, "The reliability of blood film examination for Malaria at the peripheral health unit," *Ethiopian J. Health Develop.*, vol. 17, no. 3, pp. 197–204, 2003.

[15] R. E. Coleman, K. Thimasarn, R. S. Miller, N. Maneechai, J. Sattabongkot, V. Soyseng, N. Rachaphaew, and C. Kumpitak, "Comparison of field and expert laboratory microscopy for active surveillance for asymptomatic plasmodium falciparum and plasmodium vivax in western Thailand.," *Amer. J. Tropical Med. Hygiene*, vol. 67, no. 2, pp. 141–144, Aug. 2002.

[16] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Analysis of infected blood cell images using morphological operators," *Image Vis. Comput.*, vol. 20, no. 2, pp. 133–146, Feb. 2002.

[17] M. Maitra, R. Kumar Gupta, and M. Mukherjee, "Detection and counting of red blood cells in blood cell images using Hough transform," *Int. J. Comput. Appl.*, vol. 53, no. 16, pp. 13–17, 2012.

[18] N. H. Mahmood, "Red blood cells estimation using Hough transform technique," *Signal Image Process., Int. J.*, vol. 3, no. 2, pp. 53–64, Apr. 2012.

[19] S. S. Savkare, A. S. Narote, and S. P. Narote, "Automatic blood cell segmentation using K-Mean clustering from microscopic thin blood images," in *Proc. 3rd Int. Symp. Comput. Vis. Internet (VisionNet)*, 2016, pp. 8–11.

[20] C. Ma, P. Harrison, L. Wang, and R. L. Coppel, "Automated estimation of parasitaemia of plasmodium yoelii-infected mice by digital image analysis of Giemsa-stained thin blood smears," *Malaria J.*, vol. 9, no. 1, p. 348, 2010.

[21] J. Gatc, F. Maspiyanti, D. Sarwinda, and A. M. Arymurthy, "Plasmodium parasite detection on red blood cell image for the diagnosis of malaria using double thresholding," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, Sep. 2013, pp. 381–385.

[22] J. M. Sharif, M. F. Miswan, M. A. Ngadi, M. S. H. Salam, and M. M. bin Abdul Jamil, "Red blood cell segmentation using masking and watershed algorithm: A preliminary study," in *Proc. Int. Conf. Biomed. Eng. (ICoBE)*, Feb. 2012, pp. 258–262.

[23] D. López-Puigdollers, V. Javier Traver, and F. Pla, "Recognizing white blood cells with local image descriptors," *Expert Syst. Appl.*, vol. 115, pp. 695–708, Jan. 2019.

[24] J. Yeon, J.-D. Kim, Y.-S. Kim, C.-Y. Park, and H.-J. Song, "Effective grayscale conversion method for malaria parasite detection," in *Proc. 6th Int. Conf. Multimedia, Comput. Graph. Broadcast.*, Dec. 2014, pp. 77–81.

[25] T. Markiewicz, S. Osowski, and B. Mariańska, "White blood cell automatic counting system based on support vector machine," in *Proc. 8th Int. Conf. Adapt. Natural Comput. Algorithms*. Berlin, Germany: Springer, 2007, pp. 318–326.

[26] N. E. Ross, C. J. Pritchard, D. M. Rubin, and A. G. Dusé, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Med. Biol. Eng. Comput.*, vol. 44, no. 5, pp. 427–436, May 2006.

[27] M. Habibzadeh, A. Z. Krzy ak, and T. Fevens, "White blood cell differential counts using convolutional neural networks for low resolution images," in *Proc. 12th Int. Conf. Artif. Intell. Soft Comput.* Berlin, Germany: Springer, 2013, pp. 263–274.

[28] J. A. Quinn, R. Nakasi, P. K. Mugagga, P. Byanyima, W. Lubega, and A. Andama, "Deep convolutional neural networks for microscopy-based point of care diagnostics," in *Proc. 1st Int. Conf. Mach. Learn. for Healthcare Conf. (MLHC)*, 2016, pp. 271–281.

[29] F. B. Tek, A. G. Dempster, and I. Kale, "Malaria parasite detection in peripheral blood images," in *Proc. Brit. Mach. Vis. Conf.*, 2006, pp. 347–356.

[30] G. Díaz, F. A. Gonzalez, and E. Romero, "Infected cell identification in Thin blood images based on color pixel classification: Comparison and analysis," in *Proc. 12th Ibero Amer. Congr. Pattern Recognit. (CIARP)*. Berlin, Germany: Springer, 2007, pp. 812–821.

[31] K. Pasupa, W. Sunhem, and C. K. Loo, "A hybrid approach to building face shape classifier for hairstyle recommender system," *Expert Syst. Appl.*, vol. 120, pp. 14–32, Apr. 2019.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2012, pp. 1106–1114.

[33] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017.

[34] G. Díaz, F. A. González, and E. Romero, "A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images," *J. Biomed. Informat.*, vol. 42, no. 2, pp. 296–307, Apr. 2009.

[35] K. Pasupa, S. Vatathanavaro, and S. Tungjitnob, "Convolutional neural networks based focal loss for class imbalance problem: A case study of canine red blood cells morphology classification," *J. Ambient Intell. Hum. Comput.*, Feb. 2020.

[36] K. Pasupa, S. Tungjitnob, and S. Vatathanavaro, "Semi-supervised learning with generative adversarial networks for canine red blood cells morphology classification," *Multimedia Tools Appl.*, pp. 1–18, Mar. 2020, doi: 10.1007/s11042-020-08767-z.

[37] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Jul. 2004, pp. 985–990.

[38] W. Kudisthalert and K. Pasupa, "A coefficient comparison of weighted similarity extreme learning machine for drug screening," in *Proc. 8th Int. Conf. Knowl. Smart Technol. (KST)*, Feb. 2016, pp. 43–48.

[39] W. Kudisthalert and K. Pasupa, "Clustering-based weighted extreme learning machine for classification in drug discovery process," in *Proc. 23rd Int. Conf. Neural Inf. Process. (ICONIP)*. Cham, Switzerland: Springer, 2016, pp. 441–450.

[40] K. Pasupa and W. Kudisthalert, "Virtual screening by a new clustering-based weighted similarity extreme learning machine approach," *PLoS ONE*, vol. 13, no. 4, 2018, Art. no. e0195478.

[41] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. San Diego, CA, USA: California Technical Publishing, 1997.

[42] E. R. Dougherty, *An Introduction to Morphological Image Processing* (Tutorial Texts in Optical Engineering). Bellingham, WA, USA: SPIE, 1992.

[43] J. Serra, "Introduction to mathematical morphology," *Comput. Vis., Graph., Image Process.*, vol. 35, no. 3, pp. 283–305, 1986.

[44] P. V. C. Hough, "Method and means for recognizing complex patterns," U.S. Patent 3 069 654 A, Dec. 18, 1962.

[45] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognit.*, vol. 13, no. 2, pp. 111–122, Jan. 1981.

[46] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.

[47] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Develop.*, vol. 3, no. 3, pp. 210–229, 1959, doi: 10.1147/rd.33.0210.

[48] J. R. Koza, F. H. Bennett, D. Andre, and M. A. Keane, *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*. Dordrecht, The Netherlands: Springer, 1996, pp. 151–170.

[49] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA, USA: MIT Press, 2018.

[50] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.

[51] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[52] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.

[53] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[54] C. Uzoigwe, "The human erythrocyte has developed the biconcave disc shape to optimise the flow properties of the blood in the large vessels," *Med. Hypotheses*, vol. 67, no. 5, pp. 1159–1163, Jan. 2006.

[55] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993.
[56] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed. San Diego, CA, USA: Academic., 1994, pp. 474–485.
[57] S. Rajaraman, S. K. Antani, M. Poostchi, K. Silamut, M. A. Hossain, R. J. Maude, S. Jaeger, and G. R. Thoma, "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, p. e4568, Apr. 2018.
[58] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[59] Z.-W. Yuan and J. Zhang, "Feature extraction and image retrieval based on AlexNet," *Proc. SPIE*, vol. 10033, Aug. 2016, Art. no. 100330E.
[60] B. Houwen, "Blood film preparation and staining procedures," *Clinics Lab. Med.*, vol. 22, no. 1, pp. 1–14, 2002.
[61] L. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[62] Z. Reitermanova, "Data splitting," in *Proc. 19th Annu. Conf. Doctoral Students (WDS)*, 2010, pp. 31–36.
[63] K. Pasupa and T. Lodkaew, "A new approach to automatic heat detection of cattle in video," in *Proc. 26th Int. Conf. Neural Inf. Process. (ICONIP)*. Cham, Switzerland: Springer, 2019, pp. 330–337.
[64] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, nos. 14–15, pp. 2627–2636, 1998.
[65] M. R. Spiegel, *Vector Analysis and an Introduction to Tensor Analysis*. New York, NY, USA: McGraw-Hill, 1989.
[66] S. Siegel, *Nonparametric Statistics for the Behavioral Sciences*. New York, NY, USA: McGraw-Hill, 1956.
[67] B. Liu, Y. Wei, Y. Zhang, and Q. Yang, "Deep neural networks for high dimension, low sample size data," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2287–2293.
[68] *ELM-MATLAB*. Accessed: Dec. 12, 2019. [Online]. Available: https://github.com/ExtremeLearningMachines/ELM-MATLAB-and-Online.Sequential.ELM

WASU KUDISTHALERT received the B.Sc. degree in computer science and the M.Sc. degree in information technology from the King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 2012 and 2017, respectively, where he is currently pursuing the Ph.D. degree in information technology. His research interests include machine learning, chemoinformatics, bioinformatics, and face recognition.

KITSUCHART PASUPA (Senior Member, IEEE) received the B.Eng. degree in electrical engineering from the Sirindhorn International Institute of Technology, Thammasat University, Thailand, in 2003, and the M.Sc.Eng. and Ph.D. degrees in automatic control and systems engineering from the Department of Automatic Control and Systems Engineering, The University of Sheffield, in 2004 and 2008, respectively. He was a Research Fellow of the University of Southampton and The University of Sheffield. He is currently an Associate Professor with the Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. His main research interest includes the application of machine learning techniques in real-world applications. He was a recipient of the APNNS Young Researcher Award, in 2019.

SISSADES TONGSIMA was born in Bangkok, Thailand, in 1970. He received the B.E. degree in industrial instrumentation from the King Mongkut's Institute of Technology Ladkrabang (KMITL), in 1991, and the master's and Ph.D. degrees in computer science and engineering from the University of Notre Dame, IN, USA, in 1995 and 1999, respectively.

After his graduation, he joined the National Electronics and Computer Technology Center (NECTEC), Pathum Thani, Thailand, as a Researcher, leading a team on high-performance computing research. In 2003, he applied his expertise to a new multidisciplinary research domain, bioinformatics, and moved to the National Center for Genetic Engineering and Biotechnology (BIOTEC), Pathum Thani. He was later assigned to conduct a research on discovering single nucleotide polymorphisms for Thai population (Thai SNP discovery project). In 2003, during the beginning of this project, he obtained a Post-doctoral training in SNP data acquisition and genomic variation database at the Centre National de Genotypage, Evry, France. After the completion of the SNP discovery project, he has continually published numerous research articles in both bioinformatics and computational biology. One of his highly cited article was the Science paper *Mapping Human Genetic Diversity in Asia* (2009). With his contribution in bioinformatics, he was elected to be an Executive Committee Member of the Asia–Pacific Bioinformatics Network (APBioNet) in which he has been serving for the past ten years. In 2017, he received one of the national prestigious TRF-OHEC Researcher Award with his achievement in bioinformatics research. In 2019, he was appointed to serve as the Director of the National Biobank of Thailand, which is a government funded national science and technology infrastructure aiming to conserve Thailand biological resources and sustainably utilize them via information technology and data sciences. He also serves as an Associate Editor for the *Journal of Human Genetics* (Impact factor 2.94, Nature Publishing Group).