Malaria Parasite Detection on Microscopic Blood Smear Images with Integrated Deep Learning Algorithms

Christonson Berin Jones
Computer Science and Engineering, Madurai Institute of Engineering and Technology, Tamil
*jonesberin@gmail.com

Chakravarthi Murugamani
Department of Information Technology, Bhoj Reddy Engineering College for Women,Telangana
murugamanija@gmail.com

Abstract: *Malaria is a deadly syndrome formed by the Plasmodium parasite that spreads through the bite of infected Anopheles mosquitoes. There are several drugs to cure malaria but it is difficult to detect due to inadequate equipment and technology. Microscopic check-ups of blood smear images by experts help to detect malaria-infected parasites accurately. However, manual analysis is tedious and time-consuming as the experts have to deal with many cases. This paper presents computer assisted malaria parasite detection model by classifying the blood smear image with hybrid deep learning methods that have high accuracy for classification. In the proposed approach the blood smear images are pre-processed using bilateral filtering technique in which features are extracted with the convolutional neural network. These features are selected by the improved grey-wolf optimization, and image classification is performed with the support vector machine. To evaluate the efficiency of the proposed technique, the NIH malaria dataset is utilized and the results are compared with existing approaches in terms of accuracy, F-Measure, recall, precision, and specificity. The outcome reveals that the proposed scheme is accurate and can be more helpful to pathologists for reliable parasite detection.*

Keywords: *Blood smear images, image classification, image processing, malaria, plasmodium parasite*.

Received February 2, 2021; accepted January 10, 2022
https://doi.org/10.34028/iajit/20/2/3

1. Introduction

Malaria is the most common epidemic syndrome that has affected a lot of people all over the world. It is a serious health issue, causing millions of deaths every year across many countries. According to 2019 global malaria statistics by the World Health Organization (WHO), 228 million infected cases have been recounted with an expected 405,000 deceased [28]. It is caused by a unicellular protozoan plasmodium parasite (a small living organism) which is transmitted by the female Anopheles mosquito [16]. When an infected mosquito bites a human, the parasite enters the blood and moves towards the liver for reproduction. After growing, the parasites leave the liver, travel back to the bloodstream, and infect the Red Blood Cells (RBCs). It takes four different shapes in human blood namely ring, trophozoite, schizont, and gametocyte [27]. The symptoms are typically sensed within 10 days to 4 weeks followed by infection. Most typical symptoms include headache, high fever, fatigue, muscle, and joint pain, shivering, and vomiting [15, 21]. Malaria is confirmed as a vector-borne disease that needs proper care and treatment at the premature stage itself. Initial efficient malaria diagnosis makes it both curable and preventable [25]. On the other hand, many other diseases expose similar symptoms similar to malaria. Hence, dominant diagnostic techniques are essential for accurate detection of the Plasmodium parasite in malaria infected patients [6].

Malaria diagnosis is regularly done manually by a pathologist by inspecting the blood slides of patients through a microscope. Microscopic testing is termed as the gold standard procedure because of its popularity and reasonable price [21]. The screening process is performed by keeping the blood cells under the microscope for detecting the infected cells. The blood smear masses are usually thick due to fusion with multiple blood slides [16]. When a blood cell gets infected, the cell texture varies over time and becomes difficult to distinguish between healthy and unhealthy samples [18]. Multi-view investigations of blood smear imageries offer a high chance for detection. Nevertheless, this is found to be tedious and time-consuming since pathologists have to analyze a large volume of cases [10, 19]. The growing need for inspection and limited number of pathologists have resulted in severe health issues both socially and economically. Moreover, manual diagnosis is not standardized and it is error-prone with low-resolution images which in turn lowers the accuracy of detection [16]. To overcome these disadvantages, automated microscopic blood smear image analyzing techniques are introduced in [1, 21]. It provides the following advantages:

1. It is standardized and provides more reliable information of blood films.
2. It permits more patients to be served by decreasing the workload of pathologists.
3. It can decrease the cost for diagnosis.

In the past few years, significant improvements have been made in the field of automated malaria diagnosis [5, 17, 30]. Particularly, the involvement of machine learning algorithms for automatic detection through microscopic imageries of stained blood cells can evade human error during analysis in [2, 15]. Different machine learning approaches like an artificial neural network, Naïve Bayes, k-nearest neighbor, Support Vector Machine (SVM), etc., are employed for this purpose. These techniques investigate the color features, contour dimension, and statistical attributes for grouping the RBCs that are infected with malaria [7]. Digital holography [4], phase spectroscopy/microscopy, refractive index tomography, and optical diffraction tomography [8] are some of the quantitative imaging tools utilized for capturing information about the blood cells. Even though efforts are being made for automatic diagnosis with morphological attributes, the overlapping color intensity makes it hard to categorize the infected cells. In addition, bulk cases and low-resolution imageries reduce the quality of manual analysis and are more complex to detect the illness via traditional approaches. These drawbacks can be resolved by implementing deep learning-based algorithms [26]. In this paper, an automated malaria parasite classification model is designed with the integration of hybrid approaches.

The main contributions of this paper are given below:

1. With the integration of hybrid approaches, an automated malaria parasite classification model is proposed in this paper. The proposed classification technique includes many stages such as image pre-processing, feature extraction, feature selection, and image classification.
2. To pre-process the blood smear images, a bilateral filtering technique is employed. The pre-trained convolutional neural network is used to extract the features since it extracts all the features automatically from the image without any human interference. It requires fewer parameters, training is easy and computationally more effective as compared to other techniques. A novel feature selection technique named the improved Grey-Wolf Optimization (GWO) algorithm is used to reduce the higher dimension of features. To improve the performance of GWO based feature selection, the levy flight algorithm is integrated with the GWO. In the last step, malaria parasites are classified into uninfected cells and parasitized cells using an SVM classifier.
3. The performance of the proposed automated malaria detection model has been examined on a publicly available malaria dataset and compared with existing approaches such as in terms of accuracy, F-Measure, recall, precision, and specificity. The performance of the proposed SVM classifier for the various feature extraction techniques like Visual Geometry Group-16 (VGG-16), VGG-19, GoogleLeNet, and AlexNet are compared with Naïve Bayes (NB) and K-Nearest Neighbor (KNN) classifiers. The performance of the SVM classifier is evaluated for various kernels such as sigmoid, polynomial, linear, and Radial Basis Function (RBF). The performance of the proposed feature selection technique is compared with an existing algorithm like Cuckoo Search Algorithm (CSO), Particle Swarm Optimization (PSO), and GWO in terms of convergence curve.

The remainder of this paper is structured as follows. The literature review on existing machine learning-based malaria parasite detection models is discussed in section 2. Background information about the hybrid algorithms employed in the proposed technique is briefed in section 3. The proposed method is elaborated in section 4. Results and discussions are presented in section 5. Finally, this paper is concluded in section 6.

2. Literature Review

Jan *et al*. [6] a review on automated malaria diagnosis was conducted with microscopic blood smear images. Various improvements in the domain of computer-aided diagnoses such as image normalization, segmentation, feature extraction, and classification are analyzed to address the existent challenges. It includes human error in blood slide preparation and microscope observation, noise formation during scale adjustment and color normalization, improper segmentation to distinguish RBCs from other stained objects. The authors stated the following suggestions:

1. Image acquisition from different spectral views helps to extract more meaningful details.
2. Segmentation must be limited to an extent that allows efficient image processing without contrast adjustment or over enhancement.
3. Proper counting algorithms are required to count overlapped cells.
4. Improvised classification methods can correctly classify infected cells and identify their life-stage via optimal attributes. In [25] computer-aided automatic plasmodium parasite detection was presented from microscopic blood images. In this technique, bilateral filtering is applied for noise removal and image quality enhancement. Malaria parasites inside the blood cells are detected through the combination of adaptive thresholding and morphologic image handling processes.

Experiments are conducted on NIH malaria publicly available dataset and the performance is compared with similar existing methods [3]. The detection accuracy is found to be higher than 91%. However, it wrongly classified non-parasitic cells as parasitic since some images are degraded because of high noise or poor acquisition quality. In some cases, the visuals of parasitic cells are diluted on pre-processing or adaptive thresholding resulting in missed detection.

Singlaa and Srivastava [24] a rapid, robust, and fully automatic malaria classification framework was developed with a small number of labeled datasets. It classifies different stages of malaria with a pre-trained Convolutional Neural Network (CNN) incorporated with a multi-wavelength spatial coherence microscope. This microscope can enlarge the size of the sample which is supportive for automated cell classification with deep learning algorithms. The performance is compared with numerous pre-trained CNNs and proved to outperform the existing ones with less computation complexity. Nonetheless, the absence of a segmentation approach limits its application for samples with overlapped RBCs. The blood images to detect malaria with filtered edges and classifiers are analyzed in [13]. Median filters are applied to eradicate noise and smooth the edges of images. The abnormalities present in the blood cells are identified through feature vectors such as form factor, roundness measure, shape, count of RBCs, and parasites. Finally, the SVM classifier partitions the infected blood cells from normal blood cells. The initial pre-processing and feature extraction steps are performed according to the behavior of normal and malaria-infected cells. The classification task is compared for two algorithms namely Artificial Neural Network (ANN) and SVM in which ANN outperformed the SVM classifier with 94% accuracy.

An integrated screening approach for automatic prediction and categorization of malaria-parasite infected RBCs is offered in [11]. Each blood cell present in the image is segmented via a pixel-by-pixel segmentation strategy [12]. Then, a probability map is generated for the segmented image using a pre-trained ANN classifier. This map is later post-processed to create cropped RBC image set. At last, a capsule network classifier is employed to classify the segmented cells according to the species and stages of the parasite. It has provided high accuracy and can be useful for the automatic diagnosis of any kind of pathological illness. An automated smartphone operable malaria parasite detection model working under deep learning algorithms is established in [29]. It is modeled with two processes in which an intensity-based iterative global minimum screening process is applied to find the parasitic candidates from thick smear images. Then, a customized CNN is utilized to classify the candidates as either parasite or non-parasite. This technique is evaluated in terms of different performance metrics and proved to obtain promising results. It can be more useful in areas lacking experienced parasitologists. Even though a significant number of researchers have been trained for malaria parasite detection, it is still challenging to afford highly accurate classification. Hence, major enhancements are needed to accomplish the expectation of pathologists especially to lessen the complications encountered on manual analysis.

3. Preliminaries

3.1. Image Pre-Processing

Due to the difficulties faced on the acquisition, the images may look low intense, poor, polluted with noise, and so on. These irregularities can be removed by pre-processing the original images with common filtering approaches. The main role of pre-processing is to eliminate the presence of noise and to increase the visual image quality. It is the foremost process of any image processing application since the forthcoming stages depend purely on the quality of input images. For instance, the average filter is an eminent blurring filter that suppresses different noises (like Gaussian noise, Uniform noise) in digital images. In the proposed approach, sensing the boundary of parasites in the blood cell is essential whereas blurring the low-resolution image could deteriorate it causing misinterpretation in parasite identification. Therefore, a bilateral filter that eliminates the noise but preserves the image edges is adopted in this paper. The main advantage of the bilateral filters is unlike traditional filters which consider only the spatial distance; bilateral filters take into account the range differences (i.e., intensity or color variations) as well. Bi-lateral filters are non-linear filters whose weights depend on the contents of the original image to preserve the quality of image edges. Moreover, the computation overhead of bilateral filters is lower compared to traditional filtering schemes.

3.2. Feature Extraction

Feature extraction is a major task in pre-processing that extracts the relevant data from the images. The phenomenon of representing the images in non-visual form is termed feature extraction. It is a crucial phase in several computer vision and image processing applications as it makes the conversion from pictorial to non-pictorial data illustration. The parasitic and other stained elements are flexible substances with enormous changes in size, shape, and morphologies. The color feature is beneficial but it is not enough to differentiate other stained objects and Plasmodium parasites within same as well as dissimilar species. The features that show leading variances amongst normal

and infected cells are recognized as the feature set. Many researchers have recounted both geometric and texture features for labelling the stages of malaria infection. The convolutional neural network is adopted to perform feature extraction in this paper. It is a multi-layer perceptron feed-forward neural network comprised of a convolutional layer, a pooling layer, and a fully connected layer. CNN can extract minute features from the images that are adequate for proper classification. The convolutional layer consists of a number of feature planes connected with several neurons in rectangular form. Neurons in the same feature planes share same weights which reduce the risk of overfitting between connections. The convolutional layer does convolutional operations on the spatial positions of the entire image to create feature maps. The pooling layer down samples the feature maps to reduce dimensional complexity and to enhance the generalization of the model. The fully connected layer carries out feature extraction based on the learning process. The convolutional layer and pooled layer alternately undergo forward as well as backward propagation during the training process.

3.3. Feature Selection

The key difference between feature extraction and feature selection processes is that feature extraction combines the original features and generates a new feature set where feature selection creates a subset from the original features. Feature selection is termed as dimensionality reduction technique where appropriate features are picked and inappropriate or redundant features are discarded. This will reduce the input dimension which in turn increases the performance either by lowering the learning rate and model complexity or by improving the generalization capability and prediction accuracy. The process of selecting suitable features can also lessen the measurement cost and increase the knowledge about the problem. Due to these benefits, feature selection techniques are actively employed in real-world applications typically for regression and prediction problems. On the other hand, one of the most productive areas emerging in feature selection application is the medical field, where the aim is not only limited to dimensionality reduction but also to decrease the cost involved in extracting information from images and understanding the causes behind dissimilarities in disease diagnosis among image-analyzing experts.

3.4. Classifications

The classification process of automated malaria diagnosis is mostly adopted to predict whether or not a blood cell is affected by the malarial parasite. Support vector machine is implemented in the proposed work for pathological or non-pathological cell classification. It is a supervised learning process that uses labeled training sets to train the classifier for precise disease prediction. The main goal of SVM is to generate a decision boundary called hyper-plane to segregate the features into different classes [22]. However, the decision boundary is termed as the best if it performs better classification even with new features in the future. In this work, SVM is implemented with four types of kernels such as Radial Basis Function (RBF), linear, polynomial, and sigmoid.

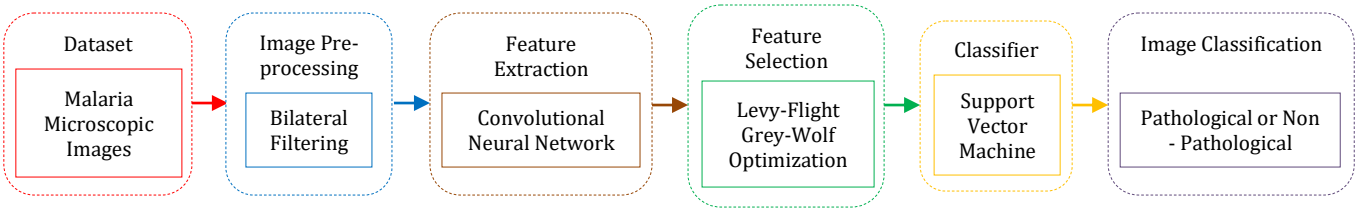


Figure 1. Block diagram of the proposed malaria disease classification mode.

4. Automated Malaria Parasite Detection

The proposed malaria disease classification model uses digital microscopic imageries that are captured from stained blood cells as the source image. The main objective is to differentiate the images into whether infected with parasites or not. The main processes of the proposed method are shown in Figure 1.

4.1. Bilateral Filtering

The main concept behind bilateral filter is the two-pixel notion in which each pixel is assumed to be near one another not only if existed in adjacent positions but also because of their similarity in the photometric distance. The use of intensity variations for edge preservation is the main benefit of a bilateral filter. The weighted addition of the pixels in a local neighborhood is estimated and its mean is utilized for changing the neighboring pixel. The outcome of the bilateral filter for a random pixel x is expressed as in Equation (1).

$$I_x(x=p)=\frac{1}{W_p}\sum_{q\in S} g_s(\|p-q\|) f_r(I_p-I_q) I_q \quad (1)$$

This is the normalized weighted mean in which p and q are pixel coordinates. S characterizes the spatial neighborhood of $I_x(x)$. g_s is a spatial Gaussian that decreases the impact of distant pixels. The range Gaussian f_r decreases the impact of pixel q if its

intensity rate varies from the intensity value of pixel *p*. I_p and I_q are intensities of pixels *p* and *q* respectively. W_p is the normalization parameter determined as given in Equation (2).

$$W_p = \sum_{q \in S} g_s(\|p - q\|) f_r(I_p - I_q) \quad (2)$$

Where *p* and *q* are the coordinates of pixels, g_s depicts the spatial kernel for smoothing the difference between coordinates. *S* denotes the spatial neighborhood of $I_x(x)$. Assume a pixel with coordinates (*i, j*) which is to be denoised with a bilateral filter and let (*m, n*) be the adjacent pixel co-ordinate. The weight to be allocated to pixel (*m, n*) to denoise the pixel at (*i, j*) is evaluated using Equation (3).

$$w(i,j,m,n) = exp\left(-\frac{(i-m)^2+(j-n)^2}{2{\sigma_d}^2} - \frac{\|I(i,j)-I(m,n)\|^2}{2{\sigma_r}^2}\right) \quad (3)$$

Here, *I*(*i, j*) and *I*(*m, n*) are intensities of pixel and σ_d and σ_r are the smoothing factors that control the function of the bilateral filter. Further, σ_d and σ_r respectively denotes the spatial and intensity domain nature of the bilateral filter. If the rates of σ_d and σ_r is increased, the significant attributes in the image get smoothened. Therefore, it is critical to deal with the initialization values of control parameters. The values of σ_r varies with the variation of noise whereas the values of σ_d do not vary much with the variation of noise. Deciding the values of σ_d and σ_r for each image is difficult as it considerably influences the functioning of the bilateral filter.

4.2. Feature Extraction with CNN

When an image is put in, the convolution layer performs processing each pixel *x* withan activation function given in Equation (4).

$$x^l = f(W^l x^l + b^l) \quad (4)$$

Where *l* denotes the number of layers, *f* indicates the activation factor, *W*, and *b* represent the weight as well as the bias of neurons. During forward propagation, a learnable convolution kernel is used to convolve the feature maps of the preceding layer. This gains a new feature map that supports the activation function given in Equation (5).

$$x_n^l = \left(\sum_{m \in M_n} X_m^{l-1} * k_{mn}^l + b_n^l\right) \quad (5)$$

The present layer *l*interpreting the preceding layer *l*-1demonstrates the first feature map of the present layer. x_n^l denotes the *n* convolution kernel for the first feature map of the present layer. The first feature map of the *n* previous layer, b_n^l is an offset value. The output of the down sampling layer is determined as shown in Equation (6).

$$x_n^l = f(\beta_n^l down(x_n^{l-1}) + b_n^l) \quad (6)$$

Where, *down* indicates the downsampling factor, b_n^l is an offset value. x_n^l denotes the *n* convolution kernel for the first feature map of the present layer. The computation overhead on feature extraction is simplified by reducing the dimension of feature maps and feature compression.

4.3. Levy-Flight Grey-Wolf Optimized Feature Selection

The mathematical modeling of GWO for the problem to be solved is initiated with the social hierarchy and hunting process of grey wolves as deliberated below.

- *Initialization*: the social hierarchy of GWO is modeled by initializing three parameters such as the fittest, the second fittest, and the third fittest respectively mentioned as alpha (α), beta (β) and delta (δ).
- *Prey Encircling*: the prey encircling procedure is mathematically formulated as given in Equations (7), and (8).

$$S_s = |a.x_p(t) - x(t)| \quad (7)$$

$$x(t+1) = x_p(t) - b.S_s \quad (8)$$

Where, S_S defines the stepsize of a greywolf. $x_p(t)$ and $x(t)$ are t^{th} iteration positions of prey and a grey wolf respectively. *a* and *b* are the coefficient vectors estimated as given in Equations (9) and (10).

$$b = 2c.r_1 - c \quad (9)$$

$$a = 2.r_2 \quad (10)$$

Where, r_1 and r_2 represents the random vectors in the range [0,1]. The factor '*c*' linearly decreases from 2 to 0 over each iteration. This factor is applied to regulate each step-size of a grey wolf.

- *Hunting*: the hunting process is directed by alpha and facilitated by beta and delta. This behavior is simulated by applying the first three best solutions (α, β *and* δ) of iteration on remaining search agents to update their locations as shown in Equations (11), (12), and (13).

$$\overrightarrow{S_{s(\alpha)}} = |\overrightarrow{a_1}.\overrightarrow{x_\alpha} - \vec{x}|, \quad \overrightarrow{S_{s(\beta)}} = |\overrightarrow{a_2}.\overrightarrow{x_\beta} - \vec{x}|, \quad \overrightarrow{S_{s(\delta)}} = |\overrightarrow{a_3}.\overrightarrow{x_\delta} - \vec{x}| \quad (11)$$

$$\overrightarrow{x_1} = \overrightarrow{x_\alpha} - \overrightarrow{b_1}.(\overrightarrow{S_{s(\alpha)}}), \quad \overrightarrow{x_2} = \overrightarrow{x_\beta} - \overrightarrow{b_2}.(\overrightarrow{S_{s(\beta)}}), \quad \overrightarrow{x_3} = \overrightarrow{x_\delta} - \overrightarrow{b_3}.(\overrightarrow{S_{s(\delta)}}) \quad (12)$$

$$\vec{x}(t+1) = \frac{\overrightarrow{x_1} + \overrightarrow{x_2} + \overrightarrow{x_3}}{3} \quad (13)$$

Where the coefficients $\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}$ can be calculated using Equation (10). The coefficients $\overrightarrow{b_1}, \overrightarrow{b_2}$ and $\overrightarrow{b_3}$ can be calculated using Equation (9). The step size of α, β *and* δ wolves are represented by $\overrightarrow{S_{s(\alpha)}}$, $\overrightarrow{S_{s(\beta)}}$ and $\overrightarrow{S_{s(\delta)}}$ respectively. The final updated position is represented by $\vec{x}(t+1)$. The position of α, β *and* δ wolves are represented by $\overrightarrow{x_1}$, $\overrightarrow{x_2}$ and $\overrightarrow{x_3}$ respectively.

- *Attacking the prey*: the grey wolf hunting behavior is stopped after catching the prey and staying in the same position without moving forward. This concept is realized by lowering the value of '*b*' in Equation (8) which depends on '*c*'. As discussed in prey encircling, the value of '*c*' is linearly reduced

from 2 to 0in search space exploitation while very few explorations might lead to local optima traps.

- *Searching for prey*: the rate of '*b*' is deliberated *either* more than +1 or less than -1 to improve the deviation behavior of GWO. This will enhance the exploration ability of GWO. Moreover, the rate of '*a*' takes a random number in all iterations denoting better exploration not only on early iterations but also the last iterations.
- *Trade-off*: exploration and exploitation are the two significant characteristics of an optimization approach. The better trade-off between these two parameters helps in attaining precise results by avoiding the trap of local optima. The step size of the search agent linearly drops with iteration which in turn is controlled by a variable '*b*'. However, GWO traps into local optima before reaching the final iterations due to poor convergence. Hence, the value '*b*' is adjusted using levy-flight that improves the capability of exploration and exploitation at the same time. Levy-flight uses the levy probability distribution function, also called the power-law function for evaluating the jump size. The mathematical expression of Levy distribution is exposed in Equation (14).

$$L(\rho,\mu,\varphi)=\begin{cases}\sqrt{\frac{\mu}{2\pi}}\exp\left[\frac{\mu}{2(\rho-\varphi)}\right]\frac{1}{(\rho-\varphi)^{\frac{3}{2}}} & if\ 0<\varphi<\infty\\ 0 & if\ \rho\le 0\end{cases} \quad (14)$$

Where ρ, μ, *and* φ are parameters that describe sample collection in the distribution, scaling factor, and position controller respectively. In the proposed model, the factor '*b*' is adjusted via levy-flights as mentioned in Equation (15). This influences the factor to take its values based on levy-flight distribution instead of linear reduction.

$$b = LEVY\ (P) * r_1 \quad (15)$$

Here, P denotes the position of wolves and r_1 is a random vector. The parameter '*b*' adjusted through levy-flight can simultaneously balance the local as well as global search behavior of GWO.

4.4. SVM based Classification

SVM is used to determine an optimal margin (hyper-plane) from a set of dissimilar planes that finely separates the features belonging to two different classes. The detection process is seen as a binary problem where +1 specifies the presence of a defect (i.e., unhealthy cell) whereas -1 specifies the absence of any defects (i.e., healthy cell). Consider a vector $x_m \in X$ of length l, where X is the time-domain and l is the number of features selected for classification. Let us consider an identifier vector $y \in$ {-1, +1}. The hyper-plane used for separation is defined as: $w^Tx + b$=0. This is formulated with the following two cases given in Equations (16), and (17).

$$w^Tx + b > 0 \quad if\ y_i = +1 \quad (16)$$

$$w^Tx + b < 0 \quad if\ y_i = -1 \quad (17)$$

Where w represents the weight vector and b is the bias weight.

5. Experimental Analysis

An automated malaria detection model must perform the tasks like image acquisition, pre-processing, feature extraction, feature selection, and classification. The proposed method is designed in this manner to differentiate the parasitic blood cells from non-parasitic blood cells. The implementation is carried out on MATLAB software running on Windows operating system. Additionally, the performance is compared to existing techniques in different forms like feature extraction, feature selection, and classification. The experimental evaluation is conducted on a publicly available microscopic image dataset. The blood smear microscopic image datasets used for analysis are publicly available at https://ceb.nlm.nih.gov/repositories/malaria-datasets/. It is a large dataset where some are shown in Figures 2, and 3. With equal instances of uninfected and parasitized cells, the dataset contains a total of 27,558 cell images. To segment and detect the red blood cells, a level-set-based algorithm was applied. The first step of image processing is pre-processing to eliminate any noise present in the images before proceeding to the next steps. In this paper, the images are filtered through bilateral filtering which produced high-quality images essential for accurate parasite detection. The parameters of the SVM is selected based on the training and testing process. The classification accuracy is considered as the selection criteria. The experiment is carried out for 25 runs with various parameter and kernel. The parameter values corresponding to the best output is selected as the parameters of the proposed model. Table 1 shows the initial parameters of the optimization algorithm. Based on the recommendation from above-mentioned existing works, these parameters are chosen.

Table 1. Parameter setting for optimizer.

Algorithm	Value
PSO	c_1,c_2=2, w_2=0.9, w_1=0.2
CSO	p_a=0.25
GWO	a=2
LFGWO	a=2, $\rho \sim U(0,1)$

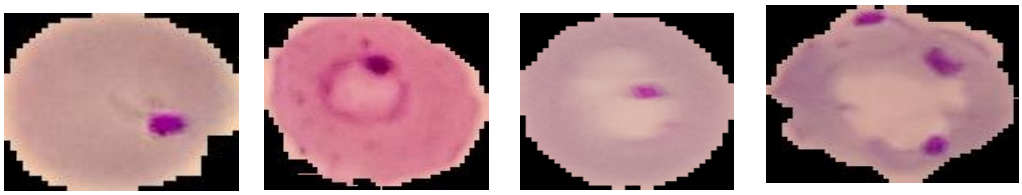


Figure 2. Sample microscopic images of red blood cells without parasites.

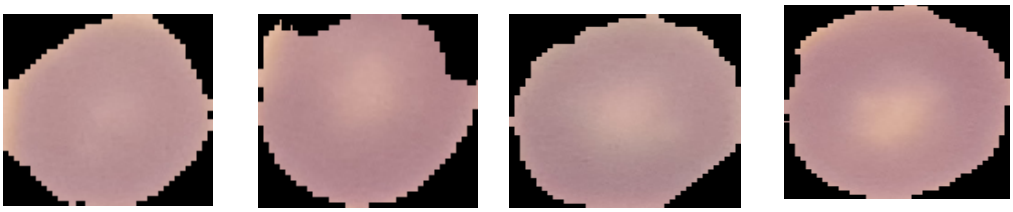


Figure 3. Sample microscopic images of red blood cells with parasites.

The comparative analysis of different feature extraction methods such as VGG-16 [23], VGG-19 [23], GoogleLeNet [9], AlexNet [31], and CNN used for malaria detection with classifiers like NB [20], KNN and SVM are listed in Table 2. It analyses the parameters such as accuracy, F-Measure, recall, precision, and specificity. These parameters can be calculated as shown by Sonibare *et al*. [25]. It can be observed that the proposed classification method has obtained significant results than existing methods in terms of all parameters. The main benefit of SVM is due to its fast learning rate with excellent generalization capability. The Receiver Operating Characteristics (ROC) curve analysis of malaria detection model for different classifiers with CNN as feature extraction is shown in Figure 4. The solid and legitimate region under ROC is generally termed as the Area Under the Curve (AUC). From the AUC of CNN-SVM, it can be noted that the proposed method produced better outcomes for positive cases rather than negative cases. The convergence curve of the proposed levy flight-based GWO feature selection technique is compared with existing approaches such as Cuckoo Search Algorithm (CSO), Particle Swarm Optimization (PSO), and GWO [14]. Moreover, levy-flight GWO reached faster convergence within 100 iterations as shown in Figure 5.

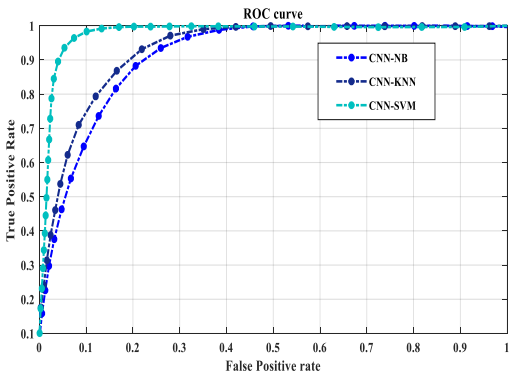


Figure 4. AUC-ROC analysis of malaria dataset for different classifiers with CNN based feature extraction.

Table 2. Comparison between feature extraction methods with different classifier models.

Feature Extraction Methods Parameters		VGG-16	VGG-19	GoogleLeNet	AlexNet	Proposed CNN
NB classifier	Accuracy	79.68	84.69	81.20	84.95	87.30
	F-Measure	73.14	79.30	80.09	74.28	81.29
	Recall	82.54	79.36	79.85	78.64	87.39
	Precision	83.21	77.94	83.75	85.21	89.36
	Specificity	82.22	85.94	81.17	86.29	89.55
KNN classifier	Accuracy	84.36	89.14	89.58	84.15	89.47
	F-Measure	81.47	86.89	85.36	81.67	84.77
	Recall	80.21	88.45	87.63	80.00	93.67
	Precision	75.34	89.62	80.97	83.68	90.78
	Specificity	81.69	87.54	86.24	80.36	90.19
SVM classifier	Accuracy	89.21	93.13	86.27	82.15	94.00
	F-Measure	87.05	91.66	83.68	77.84	93.75
	Recall	82.14	83.74	80.36	76.48	89.36
	Precision	84.47	89.95	81.56	78.04	97.00
	Specificity	88.81	92.92	86.51	85.19	97.00

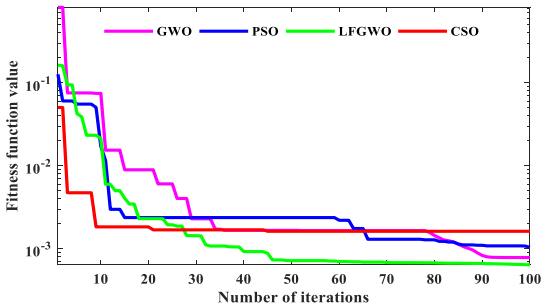


Figure 5. Convergence curve of levy-flight grey-wolf optimization.

Different parametric comparison of SVM classifier for detecting the parasite-infected cells with RBF, linear, polynomial, and sigmoid kernels is summarized in Table 3. The accuracy and F-score are respectively 94% and 93% with RBF kernel; both are less than 85% for remaining kernel functions. The precision factor is more than 80% for all the kernel functions. This demonstrates that the proposed classifier is reliable and efficient in detecting malaria parasites from microscopic blood smear images. The possible outcomes of the classifier are defined based on a confusion matrix as shown in Table 4. It includes True Positives (TP), True Negatives (TN), false positives (FP), and False Negatives (FN). The true positives

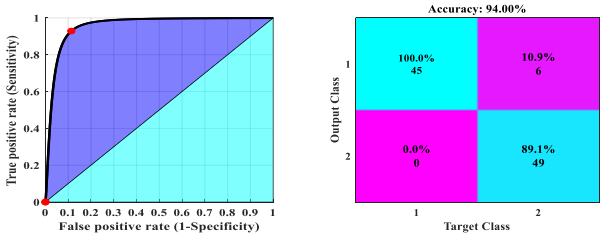
signify the infected cells detected without error and the true negatives represent the uninfected cells that are correctly detected. Similarly, false positives denote the uninfected cells wrongly detected as infected and false negatives indicate the infected cells wrongly classified as infected. The confusion matrix obtained for different kernels of the SVM classifier is given in Figure 6. Likewise, the ROC analysis of different kernels is represented in Figure 5. It can be observed that the RBF kernel of SVM has produced better results on pathological or non-pathological cell classification.

Table 3. Parameter analysis for different kernels of SVM classifier

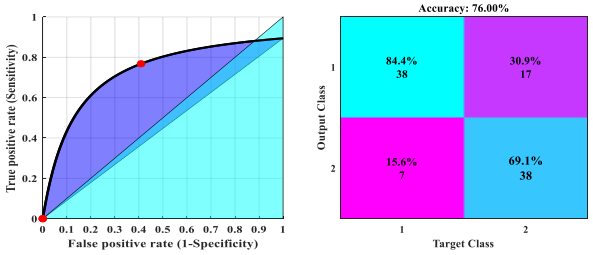
SVM kernel functions	Accuracy	Precision	F-measure	Specificity	MCC	Error	Kappa	FPR
RBF	0.9400	0.9700	0.9375	0.9700	0.8866	0.0600	0.8802	0.0300
Linear	0.7600	0.8444	0.7600	0.9444	0.5354	0.2400	0.5248	0.1556
Polynomial	0.8300	0.8667	0.8211	0.8800	0.6633	0.1700	0.6600	0.1200
Sigmoid	0.7200	0.8043	0.7255	0.7955	0.4543	0.2800	0.4453	0.2045

Table 4. The general form of the confusion matrix.

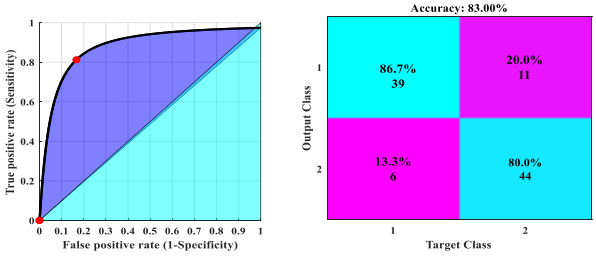
	Parasitic (infected)	Non-parasitic (uninfected)
Parasitic (infected)	TP	FN
Non-parasitic (uninfected)	FP	TN



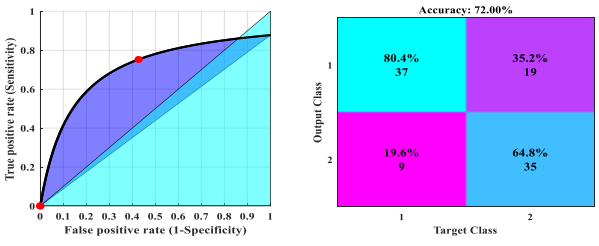
a) ROC and Confusion matrix of SVM classifier with RBF kernel function.



b) ROC and Confusion matrix of SVM classifier with Linear kernel function.



c) ROC and Confusion matrix of SVM classifier with polynomial kernel function.



d) ROC and Confusion matrix of SVM classifier with Sigmoid kernel function.

Figure 6. Analysis of SVM kernel with ROC (left column) and confusion matrix (Right column).

6. Conclusions

This paper presents an automated malaria parasite detection model from microscopic blood smear images. The bilateral filtering technique is used to pre-process the blood smear images that technique is rarely explored in the existing medical image processing approaches. The convolutional neural network adopted for feature extraction minimizes the overall risk of domain expertise with multiple processing layer. Furthermore, a novel feature selection technique based on the integration of levy-flight and the grey-wolf optimization algorithm is used to select the significant features. The performance of the proposed model is examined using a standard malaria dataset comprising a vast range of images with healthy and unhealthy blood cells. The performance of the proposed feature selection technique is compared with the existing techniques such as Cuckoo Search Optimization algorithm, Particle Swarm Optimization, and Grey Wolf Optimization. The proposed method outperformed the existing methods taken for comparison with an accuracy of more than 90%. From the experimental results, it can be known that the proposed model is effective in malaria parasite classification. In future work, the classification performance will be improved with various optimization algorithms.

References

[1] Bibin D., Nair M., and Punitha P., "Malaria Parasite Detection From Peripheral Blood Smear Images Using Deep Belief Networks," *IEEE Access*, vol. 5, pp. 9099-9108, 2017.

[2] Das D., Ghosh M., Pal M., Maiti A., and Chakraborty C., "Machine Learning Approach for Automated Screening of Malaria Parasite Using Light Microscopic Images," *Micron*, vol. 45, pp. 97-106, 2013.

[3] Fatima T. and Farid M., "Automatic Detection of Plasmodium Parasites from Microscopic Blood Images," *Journal of Parasitic Diseases*, vol. 44, no. 1, pp. 69-78, 2020.

[4] Go T., Kim J., Byeon H., and Lee S., "Machine Learning-Based in-Line Holographic Sensing of Unstained Malaria-Infected Red Blood Cells," *Journal of Biophotonics*, vol. 11, no. 9, pp. e201800101, 2018.

[5] Gomez-Aguilar J., Cordova-Fraga T., Abdeljawad T., Khan A., and Khan H., "Analysis of Fractal-Fractional Malaria Transmission Model," *Fractals*, vol. 28, no. 4, pp. 2040041, 2020.

[6] Jan Z., Khan A., Sajjad M., Muhammad K., Rho S and Mehmood I., "A Review on Automated Diagnosis of Malaria Parasite in Microscopic Blood Smears Images," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9801-9826, 2018.

[7] Jong-Dae K., Kyeong-Min N., Chan-Young P., Yu-Seop K., and Hye-Jeong S., "Automatic Detection of Malaria Parasite in Blood Images using Two Parameters," *Technology and Health Care*, vol. 24, no. 1, pp. 33-39, 2016.

[8] Kim K., Yoon H., Diez-Silva M., Dao M., Dasari R., and Park Y., "High-Resolution Three-Dimensional Imaging of Red Blood Cells Parasitized by Plasmodium Falciparum and in Situ Hemozoin Crystals Using Optical Diffraction Tomography," *Journal of Biomedical Optics*, vol. 19, no. 1, pp. 011005, 2013.

[9] Krizhevsky A., Sutskeve I., and Hinton G., "ImageNet Classification with Deep Convolutional Neural Networks Mark," *Communication of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.

[10] Kumar H. and Tolia N., "Getting in: the Structural Biology of Malaria Invasion," *PLoS pathogens*, vo1. 15, no. 9, 2019.

[11] Maity M., Jaiswal A., Gantait K., Chatterjee J., and Mukherjee A., "Quantification of Malaria Parasitaemia using Trainable Semantic Segmentation and Capsnet," *Pattern Recognition Letters*, vol. 138, pp. 88-94, 2020.

[12] Manescu P., Shaw M., Elmi M., Neary-Zajiczek L., Claveau R., Pawar V., and Oladejo B., "Expert-Level Automated Malaria Diagnosis on Routine Blood Films with Deep Neural Networks," *American Journal of Hematology*, vol. 95, no. 8, pp. 883-891, 2020.

[13] Memon M., Khanzada T.., Memon S., and Hassan S., "Blood Image Analysis to Detect Malaria Using Filtering Image Edges and Classification," *Telecommunication Computing Electronics and Control* , vol. 17, no. 1, pp. 194-201, 2019.

[14] Mirjalili S., Mirjalili S., and Lewis A., "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46-61, 2014.

[15] Park H., Rinehart M., Walzer K., Chi J., and Wax A., "Automated Detection of P. Falciparum Using Machine Learning Algorithms with Quantitative Phase Images of Unstained Cells," *PloS one*, vol. 11, no. 9, 2016.

[16] Pattanaik P., Mittal M., Khan M., and Panda S., "Malaria Detection Using Deep Residual Networks with Mobile Microscopy," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 5, pp. 1700-1705, 2020.

[17] Pillay E., Khodaiji S., Bezuidenhout B., Litshie M., and Coetzer T., "Evaluation of Automated Malaria Diagnosis using the Sysmex XN-30 Analyser in A Clinical Setting," *Malaria Journal*, vol. 18, no.1, pp. 1-14, 2019.

[18] Rehman A., Abbas N., Saba T., Mehmood Z., Mahmood T., and Ahmed K., "Microscopic Malaria Parasitemia Diagnosis and Grading on Benchmark Datasets," *Microscopy Research and Technique*, vol. 81, no. 9, pp. 1042-1058, 2018.

[19] Rosado L., Da-Costa J., Elias D., and Cardoso J., "A Review of Automatic Malaria Parasites Detection and Segmentation in Microscopic Images," *Anti-Infective Agents*, vol. 14, no. 1, pp. 11-22, 2016.

[20] Sajana T. and Narasingarao M., "Classification of Imbalanced Malaria Disease Using Naïve Bayesian Algorithm," *International Journal of Engineering and Technology*, vol. 7, no. 2.7, pp. 786-790, 2018.

[21] Shah D., Kawale K., Shah M., Randive S., and Mapari R., "Malaria Parasite Detection Using Deep Learning: (Beneficial to humankind)," *in Proceedings of the 4th International Conference on Intelligent Computing and Control Systems*, Madurai, pp. 984-988, 2020.

[22] Sharma P., Banerjee S., Tiwari D., and Patni J., C., "Machine Learning Model for Credit Card Fraud Detection-A Comparative Analysis," *The International Arab Journal of Information Technology*, vol.18, no. 6, pp. 789-796, 2021.

[23] Simonyan K. and Zisserman A., "Very Deep Convolutional Networks for Large-Scale Image Recognition," *in Proceedings of the International Conference on Learning Representations*, San Diego, pp. 1-14, 2015.

[24] Singlaa N. and Srivastava V., "Deep Learning Enabled Multi-Wavelength Spatial Coherence Microscope for the Classification of Malaria-Infected Stages with Limited Labelled Data Size," *Optics and Laser Technology*, vol. 130, pp. 106335, 2020.

[25] Sonibare O., Bello I., Olowookere S., Shabi O., and Makinde N., "Effect of Malaria Preventive Education on the Use of Long-Lasting Insecticidal Nets Among Pregnant Females in A Teaching Hospital in Osun State, South-West Nigeria," *Parasite Epidemiology and Control*, vol. 11, 2020.

[26] Torres K., Bachman C., Delahunt C., Baldeon J., Alava F., Vilela D., Ostbye T., and et al., "Automated Microscopy for Routine Malaria Diagnosis: A Field Comparison on Giemsa-Stained Blood Films in Peru," *Malaria Journal*, vol. 17, no. 1, pp. 1-11, 2018.

[27] Wood B., Bambery K., Dixon M., Tilley L.,

Nasse M., Mattsone E., and Hirschmugl C., "Diagnosing Malaria Infected Cells at The Single Cell Level Using Focal Plane Array Fourier Transform Infrared Imaging Spectroscopy," *The Analyst*, vol. 139, no. 19, pp. 4769-4774, 2014.

[28] World Health Organization, World Malaria Report, https://www.who.int/news-room/feature-stories/detail/world-malaria-report-2019, Last Visited, 2019.

[29] Yang F., Poostchi M., Yu H., Zhou Z., Silamut K., Yu J., Maude J., Jaeger S., and Antani S., "Deep Learning for Smartphone-Based Malaria Parasite Detection in Thick Blood Smears," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1427-1438, 2020.

[30] Yu H., Yang F., Rajarama S., Ersoy I., Moallem G., Poostchi M., Palaniappan K., Antani S., Maude R., and Jaeger S., "Malaria Screener: A Smartphone Application for Automated Malaria Screening," *BMC Infectious Diseases*, vol. 20, no. 1, pp. 1-8, 2020.

[31] Yuan Y. and Meng M., "Deep Learning for Polyp Recognition in Wireless Capsule Endoscopy Images," *Medical Physics*, vol. 44, no. 4, pp. 1379-1389, 2017.

Christonson Berin Jones is a Professor at Department of Computer Science and Engineering in Madurai Institute of Engineering and Technology, Sivagangai District, Tamil Nadu, India. He obtained his BE ME and PhD degree at Manonmaniam Sundaranar University, Tirunelveli, TamilNadu, India. His Ph.D research focuses on Image Processing for various applications."

Chakravarthi Murugamani received his B. Tech degree in Information Technology, from Anna University, Chennai, India in 2005; and his M.Tech degree in Information Technology from Sathyabama University, Chennai, India in 2011. He completed his Ph.D degree in Information Technology from St. Peter's University, Chennai, India in 2017. He is currently working as Professor and Head in the Department of Information Technology with the Bhoj Reddy Engineering College for Women, Hyderabad, India. He has published more than 10 International Journals along with 4 International and National conferences. He has even published 1 patent in IPR. He is also an active member in CSI and ISTE. He has received 2 times Certificate of Appreciation from IIT Madras, India for Instrumental role as SPOC for the SWAYAM- NPTEL Local Chapter for Bhoj Reddy Engineering College for Women, Hyderabad Active SPOC based on Performance and Participation of Candidates. His research and teaching interests include Computer Networks, Artificial Intelligence, Image Processing and Computer Graphics. He is an active researcher, reviewer and editor for many international journals.