

# Image classification of malaria using hybrid algorithms: convolutional neural network and method to find appropriate K for K-Nearest neighbor

Wisit Lumchanow, Sakol Udomsiri

Department of Electrical Engineering, Faculty of Engineering, Pathumwan Institute of Technology, Thailand

---

## Article Info

### Article history:

Received Jan 1, 2019

Revised Apr 6, 2019

Accepted May 10, 2019

---

### Keywords:

AlexNet

CNN

Image classification

KNN

Malaria

---

## ABSTRACT

This paper presents image classification algorithms to improve the learning rate and to comparison the classification efficiency. Using convolutional neural network (CNN) for feature extraction and method to find appropriate k for k-nearest neighbor (KNN). Medical datasets were used in the experiments to classify Plasmodium Vivax and Plasmodium Falciparum. Results of the study indicated that for Plasmodium Vivax in ring form, the appropriate k was 1 and the learning rate (LR) was 83.33%, Trophozoite (k=5, LR=91.67%), Schizont (k=1, LR=83.33%), and Gametocyte (k=1, LR=91.67%) whereas Plasmodium Falciparum in ring form (k=7, LR=91.67%), Trophozoite (k=1, LR=83.33%), Schizont (k=1, LR=91.67%) and Gametocyte (k=1, LR=100%).

Copyright © 2019 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Wisit Lumchanow,  
Department of Electrical Engineering, Faculty of Engineering,  
Pathumwan Institute of Technology,  
Bangkok, Thailand.  
Email: wisit.bru.ac@gmail.com

---

## 1. INTRODUCTION

Malaria is a disease widespread in the tropical and temperate zones, including Southeast Asia. This disease is recognized as a major public health problem of Thailand. The number of infections is highest along Thai-Myanmar and Thai-Cambodian borders [1]. Medically, malaria can be diagnosed by means of patients' blood test and microscopy where blood is placed on a slide before a test is performed to find irregularity of red blood. This is a standard test to confirm malaria infection. There are five species of malaria parasites that commonly infect humans, namely Plasmodium vivax, Plasmodium falciparum, Plasmodium malariae, Plasmodium ovale [2, 3] and Plasmodium knowlesi [4]. In Thailand, Plasmodium falciparum and Plasmodium vivax [5], are the most frequent and widely distributed causes of malaria. At present, studies to discover tools for malaria diagnosis are widely conducted, especially analysis through digital image processing system for image data classification and malaria parasite recognition by means of Gray-Level Co-Occurrence Matrix (GLCM) feature extraction [6], edge detection of malaria parasites [7], HSV color model using color histogram to construct image index in the classification process [8], threshold, grayscale, and edge detection [9]. These techniques and methods have been developed to classify image data and recognize malaria parasites up to today.

Recently, convolutional neural network (CNN) have been developed by AlexNet [10, 11], and face recognition algorithm is further developed with deep learning [12]. Route exploration of natural environment (DNN) [13], with prominent features of CNN algorithm, which can be used to extract main features of an image through filters in each convolutional layer to obtain important information of an image that can be used to effectively classify image data with k-nearest neighbor (KNN) algorithm technique [14, 15]. And from the

importance of the problems mentioned above, the researcher has an idea to develop an automatic system for reliable classification of malaria and low cost using Digital Image Processing Techniques.

In this paper is aimed at developing malaria parasite classification and recognition to enable accurate disease diagnosis with CNN for feature extraction and method to find appropriate k for KNN algorithm to improve the efficiency of detection of Plasmodium vivax, Plasmodium falciparum, to compare the test results of the proposed algorithm.

**2. RESEARCH METHOD**

This part discusses the main research method that used on this research.

**2.1. Proposed technique**

The proposed recognition and classification of malaria is made up of two types such as training and the testing, all the training malaria images are used the features extraction with CNN and KNN classification. And comparison on the efficiency and classification accuracy between the CNN-KNN and GLCM-KNN. Block diagram of classification is shown in Figure 1.

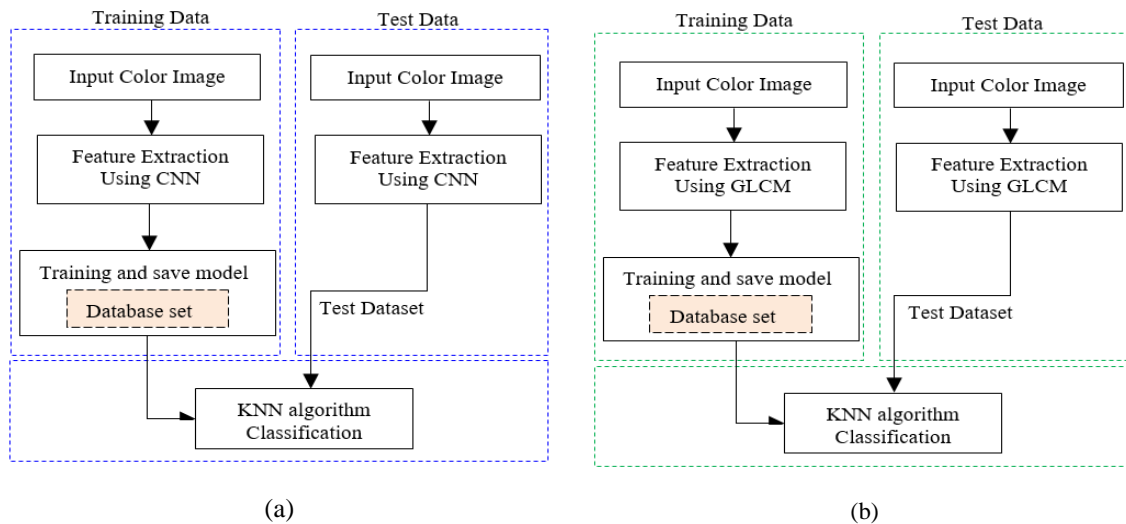


Figure 1. Block diagram of training and testing of malaria classification model, (a) Block diagram of CNN-KNN classification, (b) Block diagram of GLCM-KNN classification

**2.2. Architecture overview of the proposed system**

In this section, detailed structure of the image data classification system using CNN for feature extraction and KNN classification. Architecture of the proposed system is shown in Figure 2.

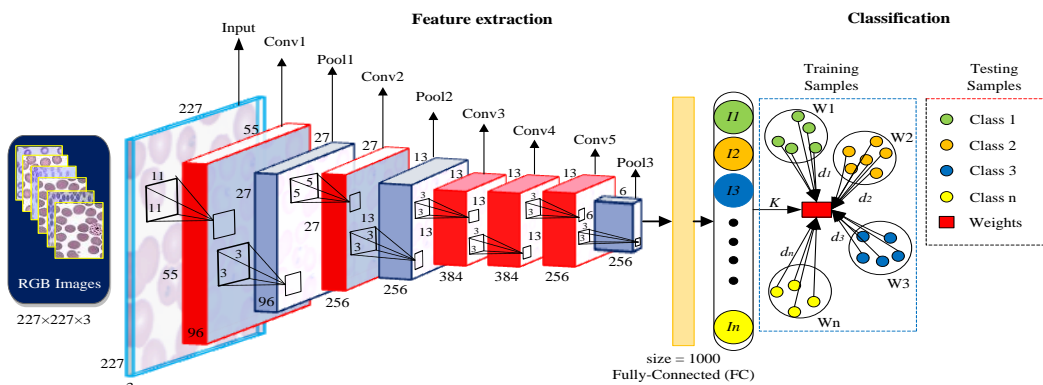


Figure 2. Structure of the CNN and KNN hybrid model

### 2.3. Convolutional neural network (CNN)

Convolutional layer [11, 16] the process to extract main features of an image through filters to obtain important information of the image. Generally, input data have pixel values of three channels, including red, blue and green. Each pixel is represented by a number to indicate color intensity. The values range from 0 to 255, representing minimum to maximum intensity. To make a black and white image, there will be only one channel, which is the channel of black, with the values of 0 (white) to 255 (pitch black). In this step, we will discuss how convolution method is used to calculate the main features of an image, gives an example of matrix of input image datasets and filter matrix. Sharpening filter matrix results are stored in smaller matrices, or feature map. Convolution performed in each layer consists of four steps as follows:

Step 1: Placing a filter on the position of input image

Step 2: Calculating the position of center of datasets from Equation (1)

$$(I \times h) \begin{bmatrix} i, j \end{bmatrix} = \sum_{u,v} h(u, v) \times I(i + u, j + v) \quad (1)$$

Where  $I$  is input matrix whose size is equal to that of the filter;

$h$  is medium matrix;

$i$  and  $j$  are positions being considered by the central position of matrix; and

$u$  and  $v$  are positions of the medium used in the calculation.

Step 3: Moving filters to different positions on the input image pixel to pixel until the last one

Step 4: Repeating Step 2 to Step 3.

Max pooling layer [11, 16]. From the process of convolution algorithm, the more convolutional layers require more time for calculation. In this regard, max pooling is an algorithm used to reduce calculation problems and increase calculation speed.

Upon completion of the feature extraction process using filters and max pooling, output will be in the form of dataset which must be pre-processing by means of re-arrangement to be in the form of one-dimension vector. From the CNN structure, a fully connected layer (FC) is generated.

### 2.4. K-nearest neighbor classification

The process of recognition of data used in the teaching is obtained from fully connected layer of CNN. This process makes the value of main features or variables of 1000 variables sent to KNN algorithm [17], which is an approach suitable for classification of medical image data. With respect to learning characteristics of KNN algorithm, no classification model is pre-established. If new image data are to be classified, they can be compared to the existing data in the manner of classification or categorization of datasets with the same or similar characteristics. The processing process in classifying image data will require determination of  $k$ , which is the number of existing data that is close to that of the data to be classified by choosing appropriate  $k$  value, which can be considered from characteristics of data classes. To obtain high classification accuracy, it may not be necessary to choose a high  $k$  value. If  $k$  is chosen to be equal to 3, the accuracy rate is high because the data have a small number of main features and classes, which makes them easier to be classified.

In finding the similarity, difference or closeness between new and old data by using a calculation principle to measure the difference with Euclidean distance, which will consider training datasets where  $i = (i_1, i_2, i_3, \dots, i_k)$  and variable  $i_k$  is an attribute of  $i$  and testing datasets where  $y = (y_1, y_2, y_3, \dots, y_k)$  and variable  $y_k$  is value of features of  $y$  in the property of distance function, which is a function frequently used in calculating Euclidean distance [14, 15, 18], which can be calculated from the following equation:

$$d(i, y) = \left[ \sum_{k=1}^n (i_k - y_k)^2 \right]^{\frac{1}{2}} \quad (2)$$

According to Equation (2),  $d(i, y)$  is Euclidean distance between  $i$  vector and  $y$  vector and values of  $n, k$  when  $n$  is the number of features of image in  $i$  vector and  $y$  vector.

## 3. RESULTS AND DISCUSSION

This section will show the efficiency of methods used to classify each type of malaria parasites and compare the efficiency of algorithms of feature extractions. Results are listed in topics as follows:

**3.1. Image datasets**

Input images were prepared by removing irrelevant backgrounds, which might cause errors to the networks. In this article, the images used in the experiments were obtained from microscope images at 100x magnification. Images with the size of 227x227 pixels used for feature extractions. With respect to the experiment procedure, data were divided into two sets: a training dataset and a testing dataset not undergoing system recognition. Examples of images of each type of malaria parasites used in the experiments are shown in Table 1.

Table 1. Sample images dataset used in the experiment.

Malaria	Stage	Images				
P. vivax	ring form					
	growing trophozoite					
	schizont					
	gametocyte					
	ring form					
P. falciparum	growing trophozoite					
	schizont					
	gametocyte					
	ring form					

**3.2. Experiment results using CNN and KNN hybrid model**

Experiments were conducted to determine the accuracy of proposed algorithms using image datasets of Plasmodium vivax and Plasmodium falciparum. Both malaria parasites are divided into four stages, namely ring form, growing trophozoite or trophozoite, schizont and gametocyte. In the experiment steps, data were divided into two sets, training and testing datasets not undergoing the recognition process, per each stage. With respect to classification algorithm test process, k value was properly adjusted for KNN classification. Results of Plasmodium vivax classification are shown in Figure 3.

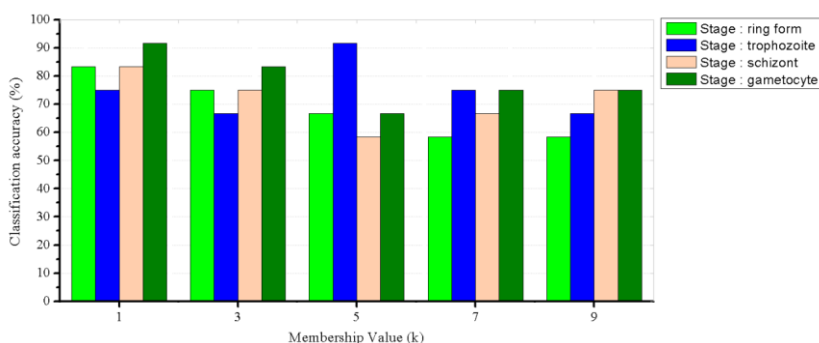


Figure 3. The accuracy testing of classification P. vivax

According to Figure 3, features of Plasmodium vivax images were extracted using the proposed algorithms with datasets of 960 images (ratio of training dataset to testing dataset was determined at 50:50). Test results of classification efficiency of Plasmodium vivax in different stages are as follows: ring form (k=1, LR=83.33%), trophozoite (k=5, LR=91.67%), schizont (k=1, LR=83.33%), and gametocyte (k=1, LR=91.67%). And the results of Plasmodium falciparum classification are shown in Figure 4.

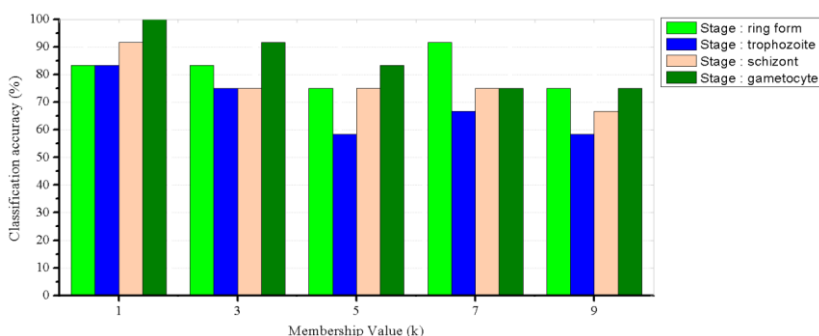


Figure 4. The accuracy testing of classification P. falciparum

According to Figure 4, main features of Plasmodium falciparum images were extracted using the proposed algorithms with datasets of 960 images (ratio of training dataset to testing dataset was determined at 50:50). Results of the classification efficiency of Plasmodium falciparum in different stages reveal that in ring form stage, the appropriate k was 7 and learning rate (LR) was 91.67%. The results for other stages are as follows: trophozoite (k=1, LR=83.33%), schizont (k=1, LR=91.67%), and gametocyte (k=1, LR=100%). It can be concluded from the experiments that major factors contributing to recognition efficiency are types of image datasets used in the experiments and level of brightness of images taken. And the results of overall accuracy of system proposed are shown in Table 2.

Table 2. The performance analysis of proposed method.

Optimal accuracy classification values								Over all accuracy of the system
Plasmodium vivax				Plasmodium falciparum				
ring form	trophozoite	schizont	gametocyte	ring form	trophozoite	schizont	gametocyte	
83.33%	91.67%	83.33%	91.67%	91.67%	83.33%	91.67%	100%	89.58%

### 3.3. Performance comparison of the proposed method with other methods

A comparison of feature extraction efficiency of malaria parasite classification using the proposed CNN for feature extraction and feature extraction using GLCM by determining k number in the KNN classification approach to have equal k value will be used in the experiments to classify Plasmodium vivax in ring form (k=1), trophozoite (k=5), schizont (k=1), gametocyte (k=1), and Plasmodium falciparum in ring form

(k=7), trophozoite (k=1), schizont (k=1), gametocyte (k=1) where k value can be determined from the result of test for appropriate k in Figure 3 and Figure 4 and datasets used in the comparative tests in which the same dataset were used to test the proposed algorithm. Comparison of the efficiency of feature extractions is shown in Figure 5. And show comparison for performance metrics accuracy. The accuracy of our proposed classification and some other methods are reported in Table 3 to prove the efficiency of our proposed system.

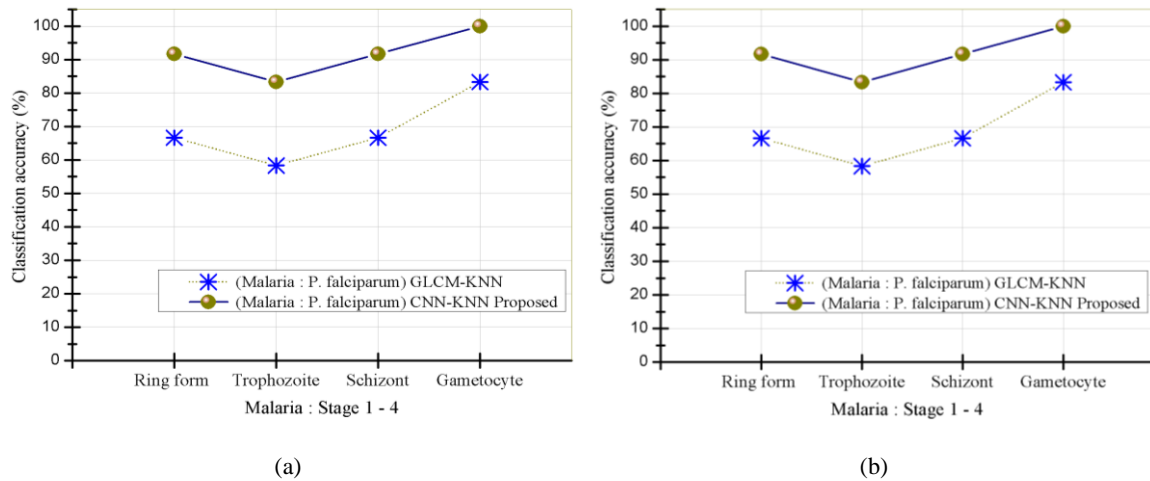


Figure 5. The comparison of classification accuracy results, (a) Malaria - Plasmodium vivax, (b) Malaria - Plasmodium falciparum

From Figure 5, comparison of Plasmodium vivax and Plasmodium falciparum classification percentages using feature extraction of CNN and feature extraction using GLCM approach, when considering the test results of malaria parasite classification in each stage, it was found that GLCM-KNN method with an average accuracy of 67.70%. It can be seen that the proposed algorithm has the higher accuracy percentage than that of the feature extraction using GLCM approach.

Table 3. The accuracy that achieved by our proposed method and other methods for malaria image classification

Related work	Techniques used for image malaria / classification	Classification Accuracy
[2]	SVM classifier (Malaria parasite detection)	86%
[3]	GLCM and KNN classifier	Not defined
[4]	Support Vector Machine (SVM)	84.2%
[19]	Levenberg-Marquardt (LM) and Bayesian Rule (BR)	87.35 for LM and 87.76 for BR
[20]	Back propagation (BP)	85%
[21]	Bayesian pixel classifier	88.5%
[22]	Neural Network Toolbox of Matlab 6.1 software package	84.85%
[23]	NN-BPFF and KNN classifier	88.06 for NN-BPFF and 83.70 for KNN
Proposed method	Feature extraction of CNN and KNN classifier	89.58%

#### 4. CONCLUSION

This article reveals the accuracy of malaria parasite classification using the CNN approach for feature extraction and KNN in order to further utilize the results obtained in developing diagnosis devices. To conduct the experiments, data were divided into two sets. The ratio of training dataset to testing dataset was determined at 50:50. With regard to the accuracy of classification of each type of malaria parasites, the results show that the learning rate is not lower than 80% and in determining appropriate k in KNN algorithm, it was found that with correct classification percentage, k was most likely to be equal to 1. It can be concluded that small number of main features of image data and classes resulted in a high rate of accuracy of malaria parasite classification.

#### ACKNOWLEDGEMENTS

This paper has been supported by Faculty of Electrical Engineering, Pathumwan Institute of Technology, Bangkok, Thailand.

## REFERENCES

- [1] P. Maneedang, D. Rongmuang, P. Thongphet and C. Sridowruang, "Lifestyle, Lack of Knowledge, Beliefs and Wrong Health Behavior Cause Malaria among Burmese Workers in Saidang Sub-District, Muang Ranong," *The Southern College Network Journal of Nursing and Public Health*, Vol. 3, No. 3, pp. 79-80, 2016.
- [2] S.K. Kumarasamy, S.H. Ong and K.S.W. Tan, "Robust contour reconstruction of red blood cells and parasites in the automated identification of the stages of malarial infection," *Machine Vision and Applications*, Vol. 22, Issue.3, pp. 461-469, 2011. doi: 10.1007/s00138-010-0284-x.
- [3] S. Sankaran, M. Pushpahaasa, R. Akshay, K. Karthikayani, B. Tech and M. Tech, "Detection and Classification of Malaria Parasites Using Digital Image Processing," *International Research Journal of Engineering and Technology*, Vol. 4, Issue. 5, pp. 87-88, 2017.
- [4] A. Nanoti, S. Jain, C. Gupta and G. Vyas, "Detection of malaria parasite species and life cycle stages using microscopic images of thin blood smear," *International Conference on Inventive Computation Technologies, 2016 IEEE International*, pp. 1-6, India, Aug 2016.
- [5] Social Research, Faculty of Science Mahidol University, "Malaria," [Online]. Available: <http://www.sc.mahidol.ac.th/usr/?p=51>. [Accessed: 9-Jun-2018].
- [6] M. I. Razzak, "Automatic Detection and Classification of Malarial Parasite," *International Journal of Biometrics and Bioinformatics*, Vol. 9, Issue. 1, pp. 7-10, 2015.
- [7] E. Komagal, K. S. kumar, and A. Vigneswaran, "Recognition and Classification of Malaria Plasmodium Diagnosis," *International Journal of Engineering Research & Technology*, Vol. 2, Issue. 1, pp. 2-3, 2013.
- [8] S. Kaewkamnerd, C. Uthaipibull, A. Intarapanich, M. Pannarut, S. Chaotheing and S. Tongshima, "An automatic device for detection and classification of malaria parasite species in thick blood film," *BMC Bioinformatics*, Vol. 13, Suppl. 17:S18, 2012. doi: 10.1186/1471-2105-13-S17-S18.
- [9] P. T. Suradkar, "Detection of Malarial Parasite in Blood Using Image Processing," *International Journal of Engineering and Innovative Technology*, Vol. 2, Issue. 10, pp. 125-126, 2013.
- [10] A. Krizhevsky, I. Sutskever and G. Hinton, "Image net classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, Vol. 25, pp. 1-9, 2012.
- [11] M.M. Saufi, M.A. Zamanhuri, N. Mohammad and Z. Ibrahim, "Deep Learning for Roman Handwritten Character Recognition," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 12, No. 2, pp. 455-460, 2018.
- [12] N.A. Binti Mat Kasim, N.H. Binti Abd Rahman, Z. Ibrahim and N.N. Abu Mangshor, "Celebrity Face Recognition using Deep Learning," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 12, No. 2, pp. 476-481, 2018.
- [13] P.A. Shyam, Y. Changju, S. Krzysztof, and K. Hyongsuk, "Accurate Natural Trail Detection Using a Combination of a Deep Neural Network and Dynamic Programming," *Sensors- Journal from MDPI*, Vol. 18, Issue. 1, pp. 1-13, 2018.
- [14] O. R. Indriani, E. J. Kusuma, C. A. Sari, E. H. Rachmawanto and D. R. Ignatius Moses Setiadi, "Tomatoes classification using K-NN based on GLCM and HSV color space," *International Conference on Innovative and Creative Information Technology, 2017 IEEE International*, pp. 1-6, Indonesia, Nov 2017.
- [15] C. Hemalatha and E. Logashanmugam, "Analysis of Different M-Band Wavelet Filters for Face Recognition using Nearest Neighbor Classifier," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 12, No. 2, pp. 826-827, 2018.
- [16] N.A. Muhammad, A.Ab. Nasir, Z. Ibrahim and N. Sabri, "Evaluation of CNN, Alexnet and GoogleNet for Fruit Recognition," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 12, No. 2, pp. 468-475, 2018.
- [17] Y.K. Zamil, S.A. Ali and M.A. Naser, "Spam image email filtering using K-NN and SVM," *International Journal of Electrical and Computer Engineering*, Vol. 9, No. 1, pp. 245-254, 2019.
- [18] P. Raktrakulthum and C. Netramai, "Vehicle Classification in Congested Traffic Based on 3D Point Cloud Using SVM and KNN," *9<sup>th</sup> International Conference on Information Technology and Electrical Engineering, 2017 IEEE International*, pp. 1-6, Phuket, Thailand, Oct 2017.
- [19] N.A. Seman, N.A. Mat Isa, L.C. Li, Z. Mohamed, U.K. Ngah and K.Z. Zamli, "Classification of Malaria Parasite Species Based on Thin Blood Smears using Multilayer Perceptron Network," *International Journal of the Computer, the Internet and Management*, Vol. 16, No. 1, pp. 46-52, 2008.
- [20] 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-ical and Biological Engineering and Computing*, Vol. 44, Issue 5, pp. 427-436, 2006. doi: 10.1007/s11517-006-0044-2.
- [21] F.B. Tek, A.G. Dempster and I. Kale, "Malaria Parasite Detection in Peripheral Blood Images," *Proceedings of the British Machine Vision Conference*, pp. 1-10, Edinburgh, UK, Sep 2006.
- [22] C.Y. Gao, H.Y. Xiong, D. Yi, G.J. Chai, X.W. Yang and L. Liu, "Study on meteorological factors-based neural network model of malaria," *Chinese Journal of Epidemiology*, Vol. 24, No. 9, pp. 831-834, 2003.
- [23] M.L. Chayadevi and G.T. Raju, "Usage of ART for Automatic Malaria Parasite Identification Based on Fractal Features," *International Journal of Video & Image Processing and Network Security*, Vol. 14, No. 04, pp. 7-15, 2014.