Microscopic peripheral malarial parasite detection and classification in blood smears using Gabor Filters and Machine learning algorithms

aravinda c, Meng Lin, Udaya Kumar Reddy k r, Amar Prabhu G

Received: date / Accepted: date

Abstract

Background: Malarial fever disease mainly caused by Plasmodium parasite that is infectious to red blood cells. Manual mode of blood cell counting is a tedious process, this leads to distressing method for diagnosis. This process’s mainly impacted on larger screening process.

Introduction: The advanced stage of technology, computer aided detection and analysis of this malarial disease, based on Gabor Filters followed by the comparison of XG-Boost classifier, Support Vector Machine and Neural Network Classifier algorithms chosen as architecture of choice for recognition and classification of these malarial blood cells.

Objective: The goal of this paper is to slow down the complexity in model discrepancy’s, and bring it to more desirable robustness and generalization, through the model development which detects and classify the parasitized and uninfected blood cells in the given sample. Roughly 13750 parasitized and 13750 unparasitized samples was taken for experiments.

Results: From the experiments the models such as S.V.M achieved 94% and XG-Boost achieved 90% neural network classifier achieved 80% , out of these S.V.M performed good results in classifying and recognizing the parasitized and uninfected blood cells to increase the accuracy in decision making.

Conclusion: The accomplishment of these M.L models, pretends these problems with less variance and obtained excellent results.

Keywords Machine Learning · Neural Network Classifier · Xtended Graident Boost · Support Vector Machine · Gabor Filters · World Health Organization

1 Introduction

Comprehensively, an projected 3.4 billion people in 92 countries are at risk of being infected with malaria and mounting disease and 1.1 billion are at high
risk. A distinct and confined area found in a currently or earlier malarious area that contains the epidemiologic and ecological factors crucial for malaria diffusion. The WHO boosts the progress of rapid and economical diagnostic tests that allow for the identification of proper treatment methods[1][2][3]. Malaria is a fatal disease caused by parasites that spread to people through the bites of infected female Anopheles mosquitoes. As per survey report in the year 2019 roughly predictable 229 million cases of malaria worldwide and the death rates hoisted at 409,000 in 2019. Small kids under 5 years are the most susceptible group affected by malaria[17][18]. The report from WHO predicts that African Region carries a extremely high share of the global malaria burden. In general microscopy testing was widely accepted, the drawback to this process was time consuming and also the expected result of this diagnosis depends on the parasitologists[4] and due to misconception diagnosis this was leading to wrong treatment. In order to overcome this drawback system, an automated system for the malaria diagnosis is an evoking research to bring into more desirable treatment such as to provide reliability, to make the quantity explicit of disease accuracy, and reduce cost effectiveness in rural areas[20][21]. Generally as per the medical experts there are 5 plasmodium classes which may be harmful from malaria disease to human beings. P.Falciparum, P.Vivax, P.Malariae, P.Ovale, P.Knowlesi. Out of these the most common two classes are P.Falciparum, P.Vivax. The P.Falciparum is severe from other classes which is causing more deaths in the society. The stages of the malarial cells are as shown in the Figure 3. Observation made from the first slide shows the P.Falciparum trophozoites and gametocytes can be observed along with the white blood cells. Next the enlarged nucleus will be compared with the rest of red blood cells in the image. The second image P.Falciparum ring stages are erected with P.Schizonts.

### 2 A study of existing authors approaches

1. **Kaewkamnerd etal**: they applied the adaptive threshold on the V-value histogram method on 20 images and achieved 60% sensitivity[11].

2. **Hanif etal**: they applied the Intensity-based contrast enhancement and threshold-based segmentation on 200 patients images and achieved the qualitative results[12].

3. **Chakrabortya etal**: they applied the Color information based morphological segmentation on 75 patients images and achieved the 90% detection and 10% false positive rate[13].

4. **Elter etal**: they applied the Histogram-based adaptive thresholding and morphological operations on denoised images on 80 patients images and achieved the 90% of detection rate[14].

5. **Quinn etal**: they applied the Feature extraction from connected components and moment features on randomized tree classifier of 2900 samples and achieved 20% of sensitivity and 90% of precision[15].
6. Toress etal, they applied Dynamic local thresholding for parasite candidate on SVM and CNN classifier on 1400 image samples and achieved 90% of sensitivity[16].

3 Methodology

In order to conduct a series of experiments, a publicly available malaria data set was used. Subsequent sections will discuss data collection, classification, augmentation and data pre-processing techniques. In a series of experiments, we learned from performance and therefore, it will be discussed in the recom-
mended model architecture section. Experimental Details and experimental settings are discussed in the training details section.

3.1 Data-sets

The datasets consists of roughly 13000 samples out of which it has been classified into two folders one is parasitized as shown in Figure:1 and uninfected as shown in Figure: 2. The Figure:1 shows the presence of Plasmodium, and Figure:2 shows may be or presence of other impurities or absence of Plasmodium.
3.2 Classification of malaria cells

In recent time’s emergent amount of studies profound to the application of computer vision and machine learning technologies to the automated diagnosis of malaria. Recently allied effort [1–3], an automated analysis method was presented in [4] for detection and enactment of red blood cells infected by the malaria parasite. To organise RBCs, three different types of ML algorithms were established for prediction accuracy and promptness as RBCs classifiers.

3.3 Image Smoothing

The cell images’ was carried out with various smoothing techniques like Gaussian noise and salt and pepper noise, compare the effect of blurring via box, Gaussian, median and bilateral filters for both noisy images as per the expected results were not promising. The 2D convolution filtering was applied with various low-pass, high pass filters in removing the noise and blurring the image. An high pass filter produced the promising results by finding the edges in an cell images. a 2x2 averaging filter kernel was applied for this cell images K=1/9.

\[
\begin{bmatrix}
1 & 1 \\
1 & 1 
\end{bmatrix}
\]

The above filtering kernel resulted as per our expectations., for every pixel, 2x2 window is centered on this pixel, then all the pixels which as coming in this window were calculated on this pixel and the result was divided by 9. This values was considered for computing the average of pixel values inside the window. This was carried out to get the filtered image as output as shown in the figure 4.to figure 7. Based on these results of parasitized images the parasitic region were mostly circular in nature and hence the circular kernel was chosen for feature extraction and recognition.

3.4 Gabor Filtration technique applied

Normally many samples are visible to naked eyes as no malarial infected cells, hence these can be used to reduce overall processing run-time. In this regards to calculate the infected cell samples, statistical analysis technique was implemented. After this infected area was noticed and threshold was performed on the color image using Gabor Filter method. The outcome of this method confirmed that noise was not only present in background as well as inside RBC’s. Later morphological series was applied to fill the holes to obtain individual samples as shown in the Figure:8 and Figure:9. The orientation of the Gabor filters information depends on accuracy. The kernels of this filter are common to the 2D field and display the important features of spatial locality and orientation. The orientation of the $\phi$ and a scale $\omega$ the Gabor wavelets
(kernels, filters) are defined as mentioned in the equation

$$\psi_{\omega \phi} = \frac{\omega}{\sqrt{2\pi}C}e^{-\omega^2/(8c^2)}(e^{i\omega} - e^{-c^2/2})$$ (1)

Observe the figure 8 infected image and figure 9 uninfected image that shows the real and imaginary parts of the Gabor kernel. Let's consider the value of $I(x+y)$ as gray value at $(x,y)$. The convolution of the sample $I$ and the Gabor kernel of the scale $\omega$ and the orientation of the $\theta$ are as mentioned in the below equation.

$$G_{\omega, \phi} = I \otimes \psi_{\omega \phi}$$ (2)

This equation results were $G_{\omega, \phi}(z)$ at pixel $z=(x,y)$ which consists of two components real and imaginary. The response of the each evenly spaced orientation...
is mentioned below.

\[ I_{\omega, \theta}(z) = R_e(G, \omega, \theta(z))^2 + I m(G, \omega, \theta(z))^2 \]  

(3)

The Figures from 10 to 13 shows the images overlaid with sub-sampled which was estimated using Gabor filters. The values considered ksize = 25*25, \( \sigma = 5, \theta = 1 \times np.pi/2, \lambda = 1 \times np.pi/4, \phi = 0.8 \).
3.5 Data-preprocessing

As we knew that the behaviour and enactments of the model completely depends on the data which is fed on the supervised learning. Hence this plays a major role in decision making. Since we used Image smoothing and gabor filter for feature extraction, we obtained a large feature vector for each image. In order to classify the data properly, we normalized each vector in the range of 0 - 255. For feature selection, chi square feature selection method was applied to select the best 80% of the features for final feature vector.

4 Classifications

4.1 Support Vector Machine

SVM is generally useful for statistical learning and determining the point location of decision boundaries which results the optimal separation of classes. The SVC classifier was used in implementing the “one-against-one” approach for multi-class classification problem where the label’s were drawn from finite set of several elements. The samples of infected parasite and uninfected samples, is 2 class which was taken as the number of classes. The decision function shape option allows to transform the results of the “one-against-one” classifiers to a decision function of shape (13000\(\text{samples}, 2\text{classes}\)). Applying each classifier to the test data vectors gives one vote to the exact class.

4.2 Neural Network Classifier

The usage of Neural nets was taken for the classification and recognition, since it consists of artificial network of functions called parameters which was able to learn all the feature of the images for analyzing the new data after receiving
one or multiple inputs as shown in the Figure 14 and the architecture as show in the Table

### 4.3 XGBoost Algorithm

As we all aware that boosting builds the model from individual 'weak learners' in an iterative way. Unlike random forest, in boosting individual models are not completely built on random subsets or data/features. Gradient boosting uses gradient descent to minimize loss function. The XGBoost algorithm was used to optimize the accuracy and speed, which uses advanced L1 and L2 regularization to prevent overfitting and fast computing. By applying this algorithm the result achieved about 90% which is as shown in the table 3

### 5 Result Analysis

To achieve the accuracy for the malarial parasite detection, a series of experiments tested using various machine learning algorithms. The S.V.M Classification Accuracy Obtained was 94% as shown in table 2 The XG-Boost
Table 2  S.V.M Classification Accuracy Obtained- 94%

|                  | Precision  | Recall   | F1        |
|------------------|------------|----------|-----------|
| Paratized-0.9259 | 0.9615     | 0.9433%  |           |
| Infected-0.9565  | 0.9166     | 0.9361%  |           |

Table 3  XG-Boost Classification Accuracy Obtained- 90%

|                  | Precision  | Recall   | F1        |
|------------------|------------|----------|-----------|
| Paratized-0.8275 | 0.9230     | 0.8727%  |           |
| Infected-0.9047  | 0.7916     | 0.8444%  |           |

classification accuracy Obtained was about 90% as shown in the table 3. The Neural Network classifier accuracy obtained was about 80% as shown in the table 1.

5.1 Experimental machine configuration

The recognition system are equipped on sever for online accessing. The CPU is Intel(R) Xeon(R) CPU E5-1410 v2 @ 2.80 GHz, RAM is 8G, and OS is Ubuntu 18.04.3 LTS

Acknowledgements  This work was supported by a Grant-in-Aid for Scientists (18K18337) from the Japan Society for the Promotion of Science (JSPS), and the Ritsumeikan University Art Research Center.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. K. S. Makhija, S. Maloney, and R. Norton, “The utility of serial blood film testing for the diagnosis of malaria,” Pathology, vol. 47, no. 1, pp. 68–70, 2015.
2. M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, “Image analysis and machine learning for detecting malaria,” Transl. Res., vol. 194, pp. 36–55, Apr. 2018.
3. Z. Liang, A. Powell, I. Ersoy, M. Poostchi, K. Silamut, K. Palaniappan, P. Guo, M. A. Hossain, A. Sameer, R. J. Maude, J. X. Huang, S. Jaeger, and G. Thoma, “CNN-based image analysis for malaria diagnosis,” in Proc. BIBM, Shenzhen, China, 2017, pp. 493–496
4. Díaz, G.; González, F.A.; Romero, E. A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. J. Biomed. Inform. 2009, 42, 296–307.
5. Ohrt, C.; Sutamihardja, M.A.; Tang, D.; Kain, K.C. Impact of microscopy error on estimates of protective efficacy in malaria-prevention trials. J. Infect. Dis. 2002, 186, 540–546.
6. Alam, M.S.; Mohon, A.N.; Mustafa, S.; Khan, W.A.; Islam, N.; Karim, M.J.; Khanum, H.; Sullivan, D.J.; Haque, R. Real-time pcr assay and rapid diagnostic tests for the diagnosis of clinically suspected malaria patients in bangladesh. Malar. J. 2011, 10, 175
7. Wongruchanalaib, C.; Barcus, M.J.; Muth, S.; Sutamihardja, A.; Wernsdorfer, W.H. A review of malaria diagnostic tools: Microscopy and rapid diagnostic test (rdt). Am. J. Trop. Med. Hyg. 2007, 77 (Suppl. 6), 119–127.
8. Rajaraman, S.; Antani, S.K.; Poostchi, M.; Silamut, K.; Hossain, M.A.; Maude, R.J.; Jaeger, S.; Thoma, G.R. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ 2018, 6, e4568.
9. Yang, D.; Subramanian, G.; Duan, J.; Gao, S.; Bai, L.; Chandramohanadas, R.; Ai, Y. A portable image-based cytometer for rapid malaria detection and quantification. PLoS ONE 2017, 12, e0179161.
10. Yunda, L.; Ramirez, A.A.; Millán, J. Automated image analysis method for p-vivax malaria parasite detection in thick film blood images. Sist. Telemática 2012, 10, 9–25
11. S. Kaewkamnerd, A. Intarapanich, M. Pannarat, S. Chaotheing, C. Uthaipibull, and S. Tongsima. “Detection and classification device for malaria parasites in thick-blood films,” in Proc. IDAACS, Prague, Czech Republic, 2011, pp. 435–438.
12. N. S. M. M. Hanif, M. Y. Mashor, and Z. Mohamed, “Image enhancement and segmentation using dark stretching technique for Plasmodium Falciparum for thick blood smear,” in Proc. CSPA, Penang, Malaysia, 2011, pp. 257–260
13. K. Chakrabortya, “A Combined Algorithm for Malaria Detection from Thick Smear Blood Slides,” J. Heal. Med. Informatics, vol. 6, no. 1, pp. 179-186, Jan. 2015.
14. M. Elter, E. Hasselmeyer, and T. Zerfass, “Detection of malaria parasites in thick blood films,” in Proc. EMBS, Boston, MA, USA, 2011, pp. 5140–5144.
15. J. A. Quinn, R. Nakasi, P. K. B. Mugagga, P. Byanyima, W. Lubega, and A. Andama, “Deep Convolutional Neural Networks for Microscopy-Based Point of Care Diagnostics,” in Proc. ICMLHC, Los Angeles, CA, USA, 2016, pp. 271–281.
16. K. Torres et al., “Automated microscopy for routine malaria diagnosis: A field comparison on Giemsa-stained blood films in Peru,” Malar. J., vol. 17, no. 1, pp. 339–50, Sept. 2018.
17. WHO, Malaria microscopy quality assurance manual, Version 2. World Health Organization, 2016.
18. S. Kaewkamnerd, A. Intarapanich, M. Pannarat, S. Chaotheing, C. Uthaipibull, and S. Tongsima. “Detection and classification device for malaria parasites in thick-blood films,” in Proc. IDAACS, Prague, Czech Republic, 2011, pp. 435–438.
19. I. K. E. Purnama, F. Z. Rahmanti, and M. H. Purnomo, “Malaria parasite identification on thick blood film using genetic programming,” Proc. ICICI-BME, Bandung, Indonesia. pp. 194–198, 2013.
20. L. Rosado, J. M. C. Da Costa, D. Elias, and J. S. Cardoso, “Automated Detection of Malaria Parasites on Thick Blood Smears via Mobile Devices,” Procedia Comput. Sci., vol. 90, no. July, pp. 138–144, Dec. 2016
21. C. Mehanian, M. Jaiswal, C. Delahunt, and C. THOMPson, “Computer-Automated Malaria Diagnosis and Quantitation Using Convolutional Neural Networks,” in Proc. ICCVW, Venice, Italy, 2017, pp. 116–125.