Detection and classification of peripheral plasmodium parasites in blood smears using filters and machine learning algorithms.*

Aravinda C.V [1,*,†], Sannidhan M S [2,†], Vaishali Bangera[3,†], Jyothi Shetty[4,†], Vijaya Shetty[5,†] and VIJAYLAXMI KOCHARI[6,‡]

[1]Nitte, deemed to be University, N.M.A.M Institute of Technology, Nitte, karkala, INDIA, 574110
[2]Nitte, deemed to be University, N.M.A.M Institute of Technology, Nitte, karkala, INDIA, 574110
[3]Nitte, deemed to be University, N.M.A.M Institute of Technology, Nitte, karkala, INDIA, 574110
[4]Nitte, deemed to be University, N.M.A.M Institute of Technology, Nitte, karkala, INDIA, 574110
[5]Nitte Meenakshi Institute of Technology, Bangalore
[6]SG Balekundri Institute of Technology, Belagavi, Karnataka, INDIA

Abstract

Malaria, a fever disease primarily caused by the infectious Plasmodium parasite that targets red blood cells, poses significant challenges in diagnosis due to the laborious manual process of blood cell counting. This limitation adversely impacts larger screening processes, necessitating the development of more efficient diagnostic methods. Leveraging advancements in technology, this paper proposes a computer-aided approach for the detection and analysis of malarial disease using Gabor Filters, followed by a comparison of three classification algorithms: XG-Boost, Support Vector Machine (SVM), and Neural Network Classifier. This study aims to reduce the complexity of model discrepancies and enhance robustness and generalization. By analyzing and classifying parasitized and uninfected blood cells in a given sample, the proposed model aims to improve the accuracy of decision-making. Experimental data comprising approximately 13,750 parasitized and 13,750 unparasitized samples were used to evaluate the models. The SVM algorithm achieved an accuracy of 94%, while XG-Boost achieved 90%, and the neural network classifier achieved 80%. Among these, SVM demonstrated the most promising results in accurately classifying and recognizing parasitized and uninfected blood cells, thereby enhancing decision-making accuracy.

Keywords

Machine Learning, Neural Network Classifier, Xtended Gradient Boost, Support Vector Machine, Gabor Filters, World Health Organization

The 5th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2023), August 28-29, 2023, Kusatsu, Japan"
*Aravinda C.V.
†All authors contributed equally.
aravinda.cv@nitte.edu.in (A. C.V); sannidhan@nitte.edu.in (. S. M. S.); vaishali.bangera@nitte.edu.in (V. Bangera); $jyothi_shetty@nitte.edu.in$ (J. Shetty); vijayashetty.s@nmit.ac.in. (V. Shetty); vijaylaxmirrao@gmail.com (V. KOCHARI)
0000-0003-4588-4105 (A. C.V)

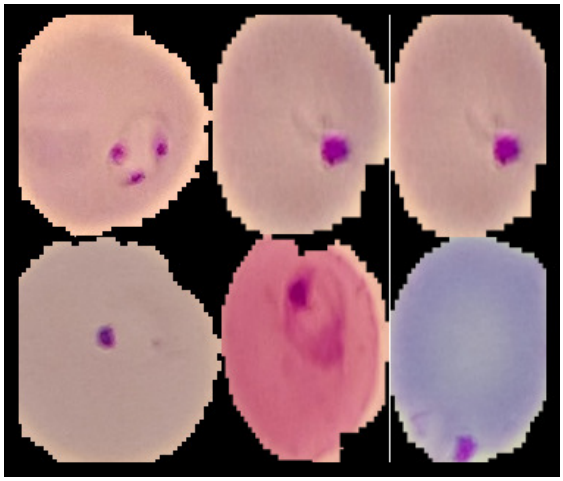

Figure 1: Malarial blood cells infected samples

1. Introduction

Malaria is a highly prevalent disease, with an estimated 3.4 billion people in 92 countries at risk of infection, of which 1.1 billion are considered to be at high risk. Specific areas within current or former malaria-endemic regions possess crucial epidemiological and ecological factors that facilitate the spread of the disease. To address this global health issue, the World Health Organization (WHO) is actively promoting the development and implementation of rapid and cost-effective diagnostic tests, which play a vital role in identifying appropriate treatment methods[1][2][3]. Malaria is caused by parasites transmitted to humans through the bites of infected female Anopheles mosquitoes. A survey conducted in 2019 estimated approximately 229 million cases of malaria worldwide, resulting in a death toll of 409,000. Young children under the age of 5 are particularly vulnerable to this disease [11][12]. The WHO report highlights that the burden of malaria is disproportionately high in the African Region. Although microscopy-based testing has been widely accepted, its time-consuming nature and dependence on skilled parapsychologists have posed limitations. Misdiagnosis based on visual interpretation has led to incorrect treatment decisions. Therefore, there is a growing need for automated systems to enhance malaria diagnosis, ensuring reliability, accurate quantification of disease, and cost-effectiveness in rural areas[14][15]. Medical experts recognize five classes of Plasmodium parasites that can cause malaria in humans: P. falciparum, P. vivax, P. malaria, P. ovale, and P. knowlesi. Among these, P. falciparum and P. vivax are the most common. P. falciparum, in particular, is associated with severe cases and higher mortality rates. The various stages of malarial cells are depicted in Figure:3. Observations from the initial slide indicate the presence of P. falciparum trophozoites and gametocytes alongside white blood cells. The enlarged nucleus of the malarial cell is then compared with the surrounding red blood cells. In the subsequent image, P. falciparum ring stages and P. schizonts can be observed.

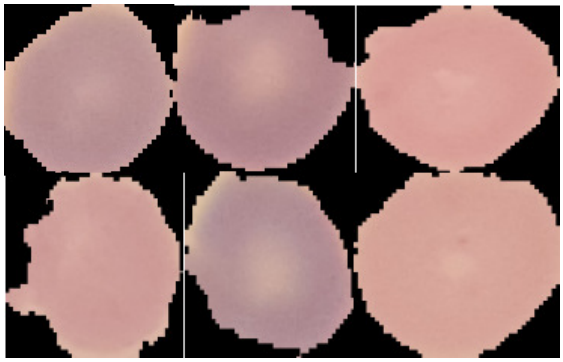

Figure 2: Malarial blood cells uninfected samples

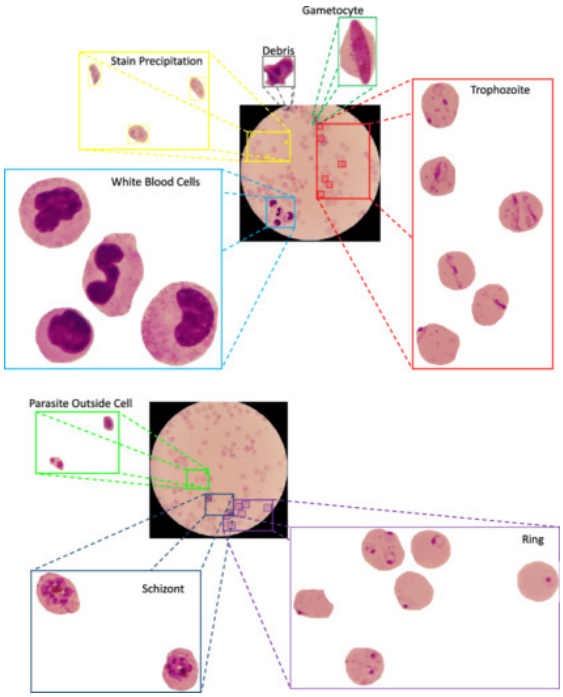

Figure 3: Stages of malaria in blood smear

2. A study of existing authors' approaches

1. **Kaewkamnerd et al :** authors worked on the flexibility of the V-value histogram method for 20 images and achieved 60% sensitivity[16].
2. **Hanif et al :** authors worked on contrast enhancement and threshold-based segmentation on 200 patient images and achieved the qualitative results[17].
3. **Chakrabortya et al :** authors worked on the Color information-based pattern segmentation on 75 patients' images and achieved the 90% detection and 10% false positive rate[18].

4. **Elter et al**they applied the Histogram-based adaptive threshold and morphological operations on denoised images on 80 patients images and achieved the 90% of detection rate[19].
5. **Quinn et al**they applied the Feature extraction from connected components and moment features on a randomized tree classifier of 2900 samples and achieved 20% of sensitivity and 90%of precision[20].
6. **Toress et al:**authors worked on pertaining to local threshold for parasite candidate on SVM and CNN classifier around 1400 image samples and achieved 90% of sensitivity[21] .
7. **Sen Li, Zeyu Du, Xiangjie Meng, and Yang Zhang** used a deep transfer graph convolutional network to build a new deep learning strategy for recognizing malaria parasites at various stages in blood smear pictures
8. **Alharbi, A. H., Aravinda, V., C, Shetty** suggested that machine-learning models be used to detect the malaria parasite in blood smear images in their study. VGG16, VGG19, ResNet50, ResNet101, DenseNet121, and DenseNet201 models were used to extract six distinct characteristics.[22]
9. **Alharbi, A. H., Aravinda, V., C, Lin, M., Ashwini, B** suggested a study that looks into using deep learning algorithms to detect a dangerous disease, malaria, for mobile healthcare solutions for patients, and to develop an efficient mobile device.[23]
10. **Md. Khayrul Bashar** proposed a supervised approach for recognizing malaria parasite stages from microscope pictures in the study. This approach combines color and texture information with a support vector machine (SVM) classifier to achieve the goal. Three texture characteristics were evaluated: an oriented pattern's histogram (HOG), a local binary pattern (LBP), and the Grey-level Co-occurrence Matrix (GLCM), as well as four color features:
11. **Park, Han Sang, and colleagues** suggested an automated analytic technique for identifying and classifying Plasmodium falciparum-infected red blood cells in the trophozoite or schizont phase of the malaria parasite Plasmodium falciparum. Quantitative phase images of unstained cells are used in this study. Various machine learning approaches such as linear discriminant classification (LDC), logistic regression (LR), and k-nearest neighbor classification.
12. **Vinayak K. Bairagi and Kshipra C. Charpe** described an automated approach for detecting malaria parasites in blood pictures. Image processing methods are utilized to diagnose and detect the stages of the malaria parasite. In blood images, factors such as statistical features and textural aspects of malaria parasites are used to diagnose parasite stages. This article compares how textural-based elements are utilized separately and how they are used in groups. The comparison is based on the accuracy, sensitivity, and specificity of the characteristics of identical photos in the database.

3. Methodology

A set of experiments was conducted using an openly available malaria data set. The upcoming sections will delve into the details of data collection, classification, augmentation, and data

pre-processing techniques. The recommended model architecture section will address the performance observed during these experiments. The training details section will comprehensively outline the entire workflow of this study.

3.1. Data-sets

The data-set comprises approximately 13,000 samples, which have been categorized into two folders: parasitized Figure: 1 and uninfected Figure.2. Figure 2 indicates the presence of Plasmodium parasites, while Figure 3 suggests the presence of other impurities or the absence of Plasmodium.

3.2. Classification of malaria cells

In recent times, there has been a growing number of studies focusing on the application of computer vision and machine learning technologies for automated malaria diagnosis. Building upon previous related research [16],[17], a recent study [19] introduced an automated analysis method to detect and identify red blood cells (RBCs) infected with the malaria parasite. To effectively classify RBCs, three distinct machine learning algorithms were implemented, aiming for accurate predictions and efficient performance as RBC classifiers.

3.3. Image Smoothing

The cell images' was carried out with various smoothing techniques like Gaussian noise and salt and pepper noise, comparing 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 and high-pass filters in removing the noise and blurring the image. A high pass filter produced promising results by finding the edges in cell images. a 2* 2 averaging filter kernel was applied for this cell image K=1/9.

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

The above filtering kernel resulted as per our expectations., for every pixel, a 2 * 2 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. These values were 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 figure 4.to figure 7. Based on these results of parasitized images the parasitic region was 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 the naked eye as no malarial infected cells, hence these can be used to reduce overall processing run-time. In this regard to calculate the infected cell samples, a statistical analysis technique was implemented. After this infected area was noticed and the threshold was performed on the color image using the Gabor Filter method.

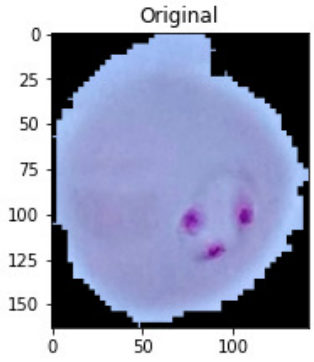

Figure 4: Infected sample

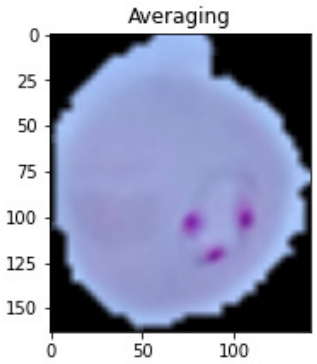

Figure 5: Infected sample

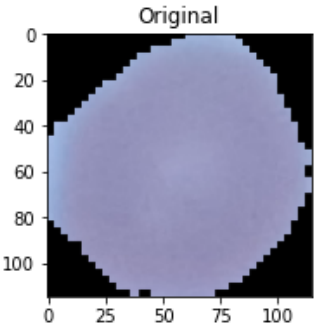

Figure 6: Uninfected sample

The outcome of this method confirmed that noise was not only present in the background as well as inside RBCs. Later morphological series was applied to fill the holes to obtain individual samples as shown in 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

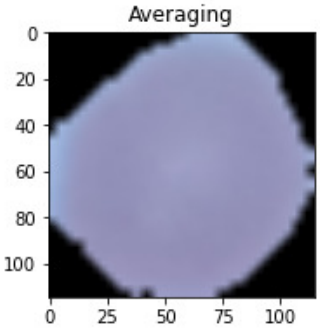

Figure 7: Uninfected sample

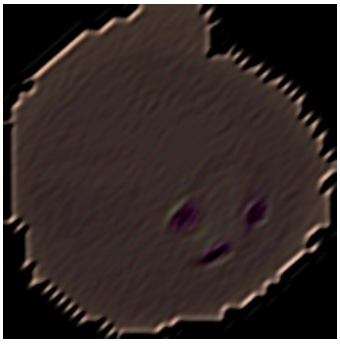

Figure 8: Gaborfilter applied

important features of spatial locality and orientation. The orientation of the ϕ and a scale ω the Gabor wavelets (kernels, filters) are defined as mentioned in the equation

$$\psi_{\omega\phi} = \frac{\omega}{\sqrt{2\pi C}} e^{-}\omega^2(4a^2 + b^2)/(8c^2)(e^{i}a\omega - e^{-}c^2/2) \tag{1}$$

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 a gray value at (x,y). The convolution of sample I and the Gabor kernel of the scale ω and the orientation of the θ are is as mentioned in the below equation.

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

This equation results were G ω, θ(z) at pixel z=(x,y)which consists of two components real and imaginary. The response of each evenly spaced orientation is mentioned below.

$$I_{\omega}\theta(z) = R_e(G, \omega, \theta(z))^2 + Im(G, \omega, \theta(z))^2 \tag{3}$$

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

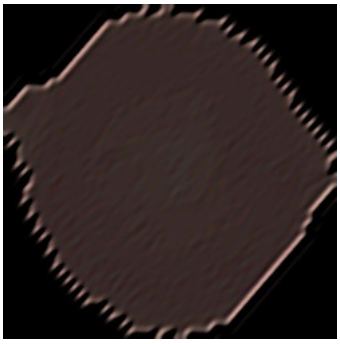

Figure 9: Gaborfilter applied

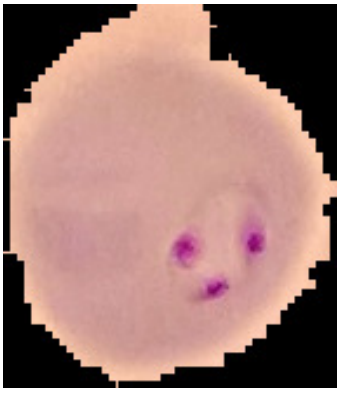

Figure 10: Infected sample

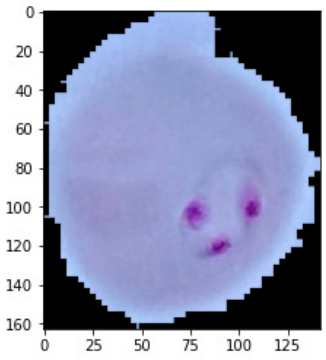

Figure 11: Image grey scale

4. Result Analysis

To achieve the accuracy for the malarial parasite detection, a series of experiments were tested using various machine learning algorithms. The S.V.M Classification Accuracy Obtained was

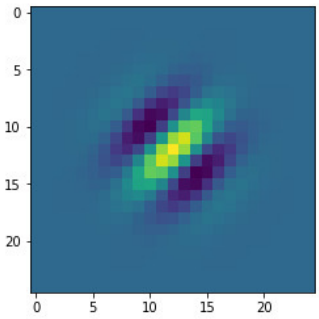

Figure 12: kernel resized

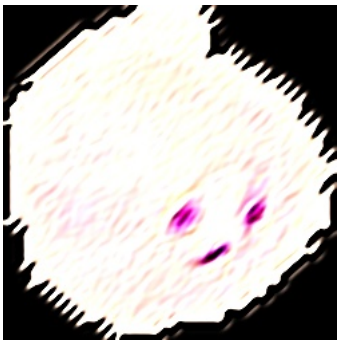

Figure 13: Output of kernel resized

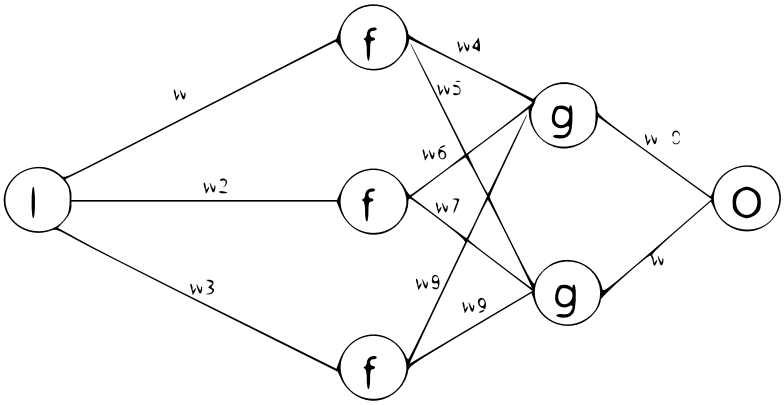

Figure 14: Simple Neural Network Architecture

94% as shown in table 2 The XG-Boost 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.

Table 1
Neural Network Structure for training

Layer (type)	Output Shape	Param
Dense	(None, 256)	48000256
Dropout	(None, 256)	0
Dense	(None, 256)	32896
Dropout	(None, 128)	0
Dense	(None, 64)	8256
Dropout	(None, 64)	0
Dense	(None, 16)	1040
Dense	(None, 1)	17
Total params: 48,042,465		
Trainable params: 48,042,465		
Non-trainable params: 0		

.

Table 2
S.V.M Classification Accuracy Obtained- 94%

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

.

Table 3
XG-Boost Classification Accuracy Obtained- 90%

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

4.1. Experimental machine configuration

The recognition system is 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

5. Comparison of models

The evaluation criteria such as precision, recall, f1-score, and accuracies of the seven ML algorithms and one ensemble algorithm that were recently explored for our multi-stage dataset can be articulated using the graph shown in figure . Precision quantifies the extent to which a model predicts a specified category. RF and NB have a higher precision score of 79%, followed by EM with 78% precision, LDA with 77% precision, KNN with 76% precision, SVM and DT with 75% precision, and LR with 55% precision. As a result, RF and NB would then predict more

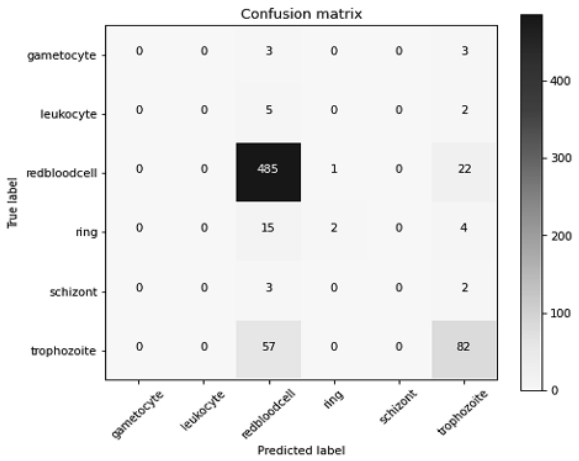

Figure 15: Randm Forest confusion Matrix

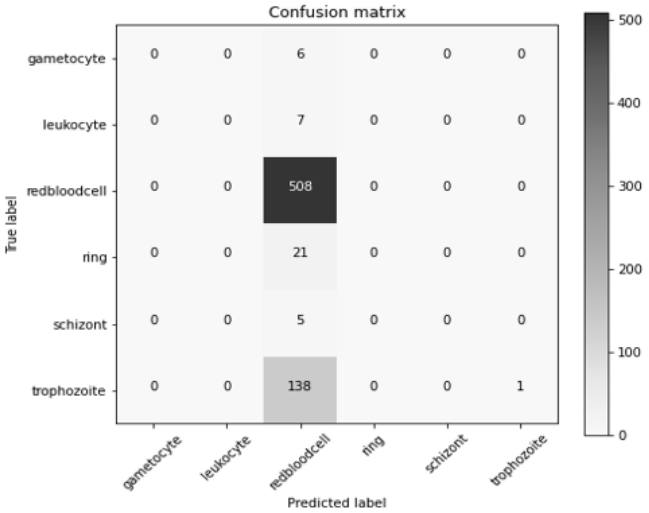

Figure 16: Support Vector confusion Matrix

relevant stage outcomes than irrelevant ones. The recall score of RF is 83%, followed by LDA and EM at 81%, KNN at 79%, SVM,LR, and DT at 74%, and NB at 53%. As a result, we can conclude that RF will correctly identify the class. The F1 score of RF is high at 80%, followed by LDA at 78%, KNN and EM at 77%, DT at 74%, LR,NB, and SVM at 63%. As a result, we can conclude that RF accurately predicts the true class of multistage malaria parasites. The confusion matrix of SVM, NN, RF, and various other models are shown in Figure 15,16,17,18 respectively.

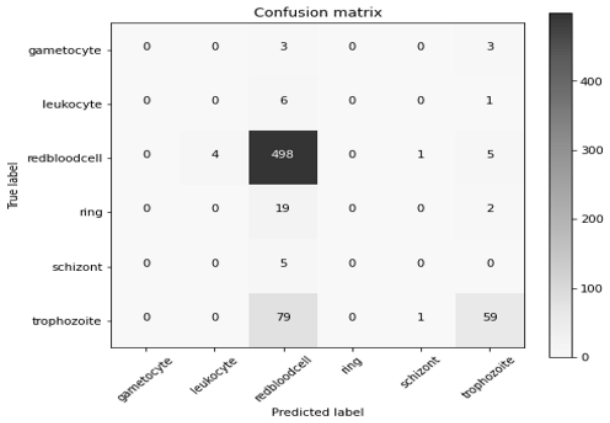

Figure 17: Neural network confusion Matrix

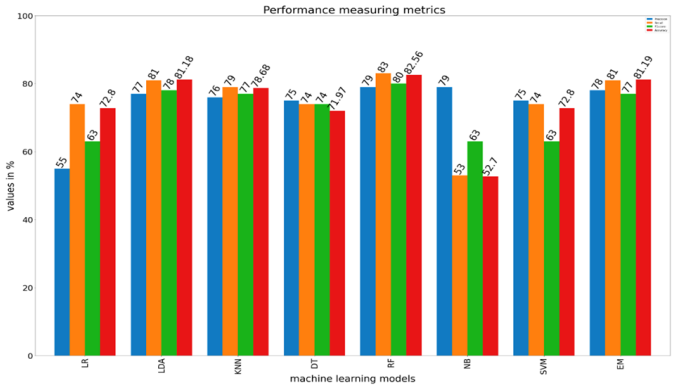

Figure 18: Various Models confusion Matrix

6. Acknowledgments

This work was supported by NITTE deemed to be University NITTE, N.M.A.M Institute of Technology Nitte, Karakala, Karnataka INDIA

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] Wongsrichanalai, 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] WHO, Malaria micropscopy quality assurance manual, Version 2. World Health Organization, 2016.

[12] 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

[13] 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.

[14] 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

[15] C. Mehanian, M. Jaiswal, C. Delahunt, and C. THOMPson, "ComputerAutomated Malaria Diagnosis and Quantitation Using Convolutional Neural Networks," in Proc. ICCVW, Venice, Italy, 2017, pp. 116–125.

[16] Alharbi, A. H., Aravinda, C. V., Lin, M., Venugopala, P. S., Reddicherla, P., & Shah, M. A. (2022). Segmentation and classification of white blood cells using the unet. Contrast Media & Molecular Imaging, 2022, 5913905. doi:10.1155/2022/5913905.

[17] 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

[18] 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.

[19] M. Elter, E. Hasslmeyer, and T. Zerfass, "Detection of malaria parasites in thick blood films," in Proc. EMBS, Boston, MA, USA, 2011, pp. 5140–5144.

[20] 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.

[21] 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.

[22] Alharbi, A. H., Aravinda, V., C, Shetty, J., Jabarulla, M. Y., Sudeepa, K. B., & Singh, S. K. (2022). Computational models-based detection of peripheral malarial parasites in blood smears. Contrast Media& Molecular Imaging, 2022, 9171343. doi:10.1155/2022/917134

[23] Alharbi, A. H., Aravinda, V., C, Lin, M., Ashwini, B., Jabarulla, M. Y., & Shah, M. A. (2022). Detection of peripheral malarial parasites in blood smears using deep learning models. Computational Intelligence and Neuroscience, 2022, 3922763. doi:10.1155/2022/3922763

[24] 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.