Plasmodium species malaria detection and comparison using the C means clustering algorithm and watershed algorithm

Sudharsan D[1], Indira K.P[2*]

[1]Research scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.

[2*]Project guide, Corresponding author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105

ABSTRACT

Aim: The aim of this project is to detect the presence of plasmodium species using modern algorithms, and comparing the peak signal to noise ratio (psnr) between Innovative clustering algorithm and the watershed algorithm. **Materials and Methods:** The sample images are taken from the Kaggles website. In accordance with total sample size calculated clinicalc.com by keeping alpha error threshold value of 0.10, enrollment ratio as 0.1, 95% confidence interval, and G power at 80%, samples were considered as N = 10 for the C-means clustering algorithm and N=10 for the watershed algorithm. PSNR was calculated using MATLAB programming with a standard dataset. **Results:** The PSNR comparison is done by an independent sample t-test using SPSS software. There seems to be a significant difference between the C-means clustering algorithm and the watershed algorithm. The PSNR of the C-means clustering algorithm is 29.3464% ($p = 0.010$) better than the Watershed algorithm (PSNR=0.2109). The SSIM of C means clustering is 4.8954, showing better results compared to the Watershed algorithm at 0.1484. **Conclusion:** C means clustering algorithm values were found to show a better PSNR than watershed algorithm for the detection of malaria.

Keywords: Malaria, Plasmodium Species, Segmentation, Innovative Clustering algorithm, Watershed algorithm, MATLAB Programming

INTRODUCTION

The project helps in detecting the presence of plasmodium species (malaria) at an early stage using modern algorithms like Innovative simplified C means clustering algorithm and watershed Algorithm . Malaria is a severe infectious disease that is spread across many countries. Malaria affects millions of people every year(Devi, Sheikh, and Laskar 2016)A proper diagnosis of malaria patients is the first and most important step (Abdurahman, Fante, and Aliy 2021). The main goal of this study is to show that, using modern algorithms, early detection of malaria will reduce the damage risk for the liver (Eliza and K. 2019) . The watershed approach is often used in therapeutic image segmentation, such as in leukemia, core, or granular form (Talari and Gholizadeh 2016). This study can be implemented in healthcare, diagnosis centers and hospitals (Dhivya et al. 2016)(Salamah et al. 2018).

About 158 Google Scholar and 39 Science Direct papers were examined in relation to this study, which has been carried out in recent years. For diagnosis, algorithm based segmentation of plasmodium species malaria-infected

blood cells in microscopic pictures is proposed . The infected blood samples were extracted using a clustering approach. This method was used to refine the final edge-based segmentation of the collected pictures . The Watershed algorithm provides better identification, especially in overlapping conditions and automation watershed is a well known and effective tool in medical image segmentation. Previously, some researchers used this technique to make diagnoses. Therefore, one of the most effective methods for segmenting images is through Innovative watershed algorithms. The study, on the other hand, is concerned with the detection of malaria using watershed and C means clustering algorithms. The malaria parasites attack the liver which will be travelling through the human bloodstream (Sampathila, Shet, and Basu 2018). When the parasites mature, they leave the liver and infect red blood cells, which is the most common form of malarial fever. This research describes an effective methodology to study any fundus using algorithms for the detection of malaria. This research encoded with MATLAB coding, which can be considered the easiest and simplest way to detect malaria, was compared to the previous study mentioned above algorithm, which is considered to be more accurate in determining the presence of malaria . This research paper was considered the best one for detecting Plasmodium species malaria compared to the other research work Previously our team has a rich experience in working on various research projects across multiple disciplines(Balusamy et al. 2020; Arvind and Jain 2021; Zhao et al. 2020; Hani et al. 2020)

In some cases, highly required treatment at a later stage is not possible. Therefore, it is a major key point that motivates us to work on this project to detect the presence of plasmodium species malaria at an early stage. The authors were experts in the fields of machine learning algorithms and MATLAB programming and were able to conduct biomedical research comparing the threshold algorithm and the Innovative Simplified C Means clustering algorithm. The major goal is to diagnose malaria early with a higher PSNR.

MATERIALS AND METHODS

This research is being carried out at the Saveetha School of Engineering, Chennai. This study has no ethical implications. The Threshold algorithm and the C-means clustering algorithm are the two groups involved in Matlab programming. Each group has a sample size of 20. Using clinalcalc.com, the sample size was calculated by using the parameter values from prior iterations (Saw 2013). The pretest G power is set at 80%, the threshold is set at 0.010, and the confidence interval is set at 95%.

Sample preparation using algorithms

Sample preparation for group 2, which consists of 20 samples, for the two processes was completed. The dataset's input images were first rescaled to 630 x 270 pixels. Following that, the C-Means clustering algorithms and watershed algorithms are used to extract and classify features (Dey 2019). MS Excel is used to export the estimated sample values for additional data methods. Instead of the predicted label of the testing image, the algorithm is trained with features from all of the images and during testing, a total

label of obtained features is expected rather than the predicted label of the testing image. If the image of the majority of features matched the expected image, the recognition is successful.

Testing setup and testing procedure

All of the experiments were conducted on the MATLAB software 2018 version with all the add-ons required for thorough training and testing. A sample of low-resolution photos was provided as an input for the testing procedure. Scaling was done in the preprocessing step to resize the photos to 600 by 450 pixels. It then recognises the malaria image for feature extraction; the output has 433 dimensions and contains more feature information, allowing for higher retrieval performance. Finally, the picture of malaria has been recognised. The picture enhancement PSNR is calculated using the Innovative Clustering algorithm and the Watershed algorithm. The SPSS IBM application uses the sample value saved in MS-Excel for statistical analysis.

STATISTICAL ANALYSIS

To validate the results of both the algorithms statistical analysis was done using IBM SPSS software. As the two algorithms are independent of each other , an independent sample t test was performed for the independent variable PSNR. There are no dependent variables.

RESULTS

In this research, detecting malaria, the C-means clustering algorithm was found to be better than the Watershed algorithm, giving a higher PSNR. The value of PSNR given by the C-means clustering algorithm is 29.3464 and that given by the Watershed algorithm is 0.2099. In Figure 1, it is observed that the mean PSNR is higher for the C-means clustering algorithm than for the Watershed algorithm.

Table 1 represents the PSNR of the sample tests collected using the Watershed algorithm and the C-means algorithm. The PSNR of the sample using the Innovative Clustering algorithm and the Watershed algorithm is given in table 1. The clear statistics in Table 2 illustrate that the C-means clustering algorithm has a lower error rate than the Watershed algorithm. Figure 2a represents the C-mean clustering picture for the given sample picture of malaria. Figure 2b represents the Watershed algorithm picture for the given sample image of malaria. The clustering algorithm produces better results compared to the watershed algorithm's mean.

To analyze the PSNR of malaria, the C-mean Clustering algorithm and the Watershed algorithm are used, as well as the independent sample t-test. There is a statistically significant difference ($p=0.010$, p 0.05) in both the methods using independent test t-tests as shown in Table 3. This methodology proposes that malaria can be identified early. The results show that the Innovative Clustering algorithm can be used to detect malaria at an early stage in comparison with the Watershed algorithm.

DISCUSSION

In this research work, detecting malaria using an Innovative clustering algorithm has the highest PSNR (29.3464) in comparison to the Watershed algorithm (0.2099). There appears to be a slight increase in the significant difference, but not in the statistically significant (Ahmad 2016). The proposed algorithm's average

segmentation accuracy was 95.2 percent, according to the experimental findings. Further investigation revealed that the malaria parasite's in blood cells were successfully removed, indicating that the approach is suitable for application in the detection of malaria parasites (Shah et al. 2016). The result of sensitivity, specificity, and accuracy should be higher to represent a good segmented image. It is shown that the proposed method achieved better performance regarding specificity (99.62%) and accuracy (98.52%) (Rajinikanth, Raja, and Dey 2020). However, regarding sensitivity, it is slightly lower (68.72%) compared to the other methods. Logically, the specificity will be high for all malaria parasite tests (Nasir et al. 2012). Thus, when compared to the previous study mentioned above, which is considered to be more accurate in determining the presence of malaria.

A few of the factors that are affecting this study might be due to the color contrast, where subjective image consistency is basic for human recognition, pixel size, perspective ratio of the picture contrast changes depending upon the medium and picture brightness. The aspect ratio and measure of the image are considered to be some of the most vital parameters.

Although the above performed algorithm has a few advantages over the other algorithms in detecting malaria and separating the other species of malarial diseases, its limitations are caused by a certain factor's inefficient real-time calculation, which can be considered more helpful in identifying malaria and modifications in the algorithm to get better PSNR.

We will study the use of detecting malaria in the healthcare area and enhancing the peak signal to noise ratio (PSNR) for diseased images in the near future. As a result, this project will have a bright future in this area, where manual labour can be simplified and reduced, and computerized production can be changed over at a low price. A much better real-time dataset and application using different machine learning and deep learning methods, such as the watershed algorithm and the c-means clustering algorithm, may produce better outcomes.

CONCLUSION

In the study of malaria expectations, the C mean clustering algorithm, which works with Novel Matlab programming and has a PSNR of 0.2099 and an SSIM of 0.1484, outperforms the Watershed algorithm, which has a PSNR of 29.3464 and an SSIM of 4.8954. Thus, it might be used in hospitals and diagnostic centers.

DECLARATION

Conflict of Interest

In this manuscript, there are no conflicts of interest.

Authors Contribution

Author SD was involved in data collection, data analysis, and manuscript writing. Author IKP was involved in conceptualisation, data validation and critical review of manuscript.

Acknowledgement

We would like to express our special thanks of gratitude to the management, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences(Formerly known as Saveetha University) for providing me with necessary infrastructure that was required in completing the project.

Funding:

We thank the following organization for providing financial support that enabled us to complete the study.

1. Axia clinics Pvt.Ltd,Hyderabad.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

1. Abdurahman, Fetulhak, Kinde Anlay Fante, and Mohammed Aliy. 2021. "Malaria Parasite Detection in Thick Blood Smear Microscopic Images Using Modified YOLOV3 and YOLOV4 Models." *BMC Bioinformatics*. https://doi.org/10.1186/s12859-021-04036-4.
2. 6040727.
3. Devi, Salam Shuleenda, Shah Alam Sheikh, and Rabul Hussain Laskar. 2016. "Erythrocyte Features for Malaria Parasite Detection in Microscopic Images of Thin Blood Smear: A Review." *International Journal of Interactive Multimedia and Artificial Intelligence*. https://doi.org/10.9781/ijimai.2016.426.
4. Dey, Nilanjan. 2019. *Classification Techniques for Medical Image Analysis and Computer Aided Diagnosis*. Academic Press.
5. Dhivya, R., Assistant Professor/ECE, Kingston Engineering College, and Vellore. 2016. "A Paper On Edge Detection In Images Using Fuzzy K Means Clustering Approach." *International Journal Of Engineering And Computer Science*. https://doi.org/10.18535/ijecs/v5i11.23.
6. Eliza, Thenu, and Sreekumar K. 2019. "Malaria Parasite Detection from Microscopic Blood Smear Image: A Literature Survey." *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2019918741.
7. Hani, Umme, Mohamed Rahamathulla, Riyaz Ali Osmani, Honnavalli Yogish Kumar, Deeparani Urolagin, Mohammad Yousuf Ansari, Kamal Pandey, Keerthana Devi, and Sabina Yasmin. 2020. "Recent Advances in Novel Drug Delivery Systems and Approaches for Management of Breast Cancer: A Comprehensive Review." *Journal of Drug Delivery Science and Technology* 56 (April): 101505.
8. Nasir, A. S. Abdul, A. S. Abdul Nasir, M. Y. Mashor, and Z. Mohamed. 2012. "Segmentation Based Approach for Detection of Malaria Parasites Using Moving K-Means Clustering." *2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences*. https://doi.org/10.1109/iecbes.2012.6498073.
9. Rajinikanth, Venkatesan, Nadaradjane Sri Madhava Raja, and Nilanjan Dey. 2020. *A Beginner's Guide to Multilevel Image Thresholding*. CRC Press.
10. Salamah, Umi, Riyanarto Sarno, Agus Zainal Arifin, Anto Satriyo Nugroho, Ismail Eko Prayitno Rozi, and Puji Budi Setia Asih. 2018. "Incorporating Index of Fuzziness and Adaptive Thresholding for Image Segmentation." *International Journal of Electrical and Computer Engineering (IJECE)*. https://doi.org/10.11591/ijece.v8i4.pp2406-2418.
11. Sampathila, Niranjana, Nagaraja Shet,

and Akash Basu. 2018. "Computational Approach for Diagnosis of Malaria through Classification of Malaria Parasite from Microscopic Image of Blood Smear." *Biomedical Research*. https://doi.org/10.4066/biomedicalresearch.29-18-970.

12. Saw, Hui Ann. 2013. *Analysis of Microscopic Blood Samples for Detecting Malaria*.
13. Shah, M. I., S. Mishra, M. Sarkar, and C. Rout. 2016. "Automatic Detection and Classification of Tuberculosis Bacilli from ZN-Stained Sputum Smear Images Using Watershed Segmentation." *International Conference on Signal Processing (ICSP 2016)*. https://doi.org/10.1049/cp.2016.1459.
14. Talari, Arash, and Roya Gholizadeh. 2016. "Hydrological Analysis and Ability Assessing of Flood Gating of Karaj Watershed According to the Form of the Watershed." *Open Journal of Geology*. https://doi.org/10.4236/ojg.2016.66029
15. Zhao, Yuanzheng, Minyan Dang, Wenzhi Zhang, Yan Lei, Thiyagarajan Ramesh, Vishnu Priya Veeraraghavan, and Xunyao Hou. 2020. "Neuroprotective Effects of Syringic Acid against Aluminium Chloride Induced Oxidative Stress Mediated Neuroinflammation in Rat Model of Alzheimer's Disease." *Journal of Functional Foods* 71 (August): 104009.

TABLES AND FIGURES

Table 1. PSNR of the sample using Innovative simplified C means clustering algorithm and Watershed algorithm.Table comparison of PSNR for malaria using two algorithms.

Samples	C means clustering algorithm (PSNR)	Watershed algorithm (PSNR)
1	29.2113	4.6923
2	31.3002	4.8030
3	27.8449	4.4424
4	28.7028	5.0703
5	31.1175	5.0703
6	29.7820	4.5101
7	27.7120	4.4705
8	26.2798	4.4897

9	29.1229	4.7727
10	32.3927	6.0333

Table 2. Comparison of PSNR and mean using the Innovative clustering algorithm and the Watershed algorithm. group statistics comparison of PSNR for malaria using the Watershed algorithm and the C means clustering algorithm is done. C means clustering has a higher mean value compared to the Watershed algorithm. C means clustering is 29.3464 and the Watershed algorithm is 0.2099

	Group	**p**	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
PSNR	C means clustering	1.00	10	29.3464	1.86317	0.58919
	Watershed	1.00	10	0.2099	0.01441	0.00456
SSIM	C means clustering	2.00	10	4.8954	0.54730	0.17307
	Watershed	2.00	10	0.1484	0.04780	0.01512

Table 3. In orderto analyze the PSNR of malaria, the C-mean clustering algorithm and the Watershed algorithm were used, as well as the independent sample t-test. There appears to be a statistically significant difference ($p < 0.05$) in both methods.

		F	**Sig**	**T**	**df**	**One tailed p**	**Two tailed p**	**Mean diff**	**Std Error Diff**	**Lower**	**Upper**
C means clustering	Equal variances assumed	8.256	0.01	39.817	18	<0.001	<0.001	24.451	0.614	23.160	25.741
	Equal variances not assumed			39.817	10.54	<0.001	<0.001	24.451	0.614	23.092	25.809
Watershed	Equal variances	4.211	0.55	3.896	18	<0.001	0.001	0.061	0.015	0.283	0.094

	assumed										
	Equal variances not assumed			3.986	10.62	0.001	0.003	0.061	0.015	0.026	0.096

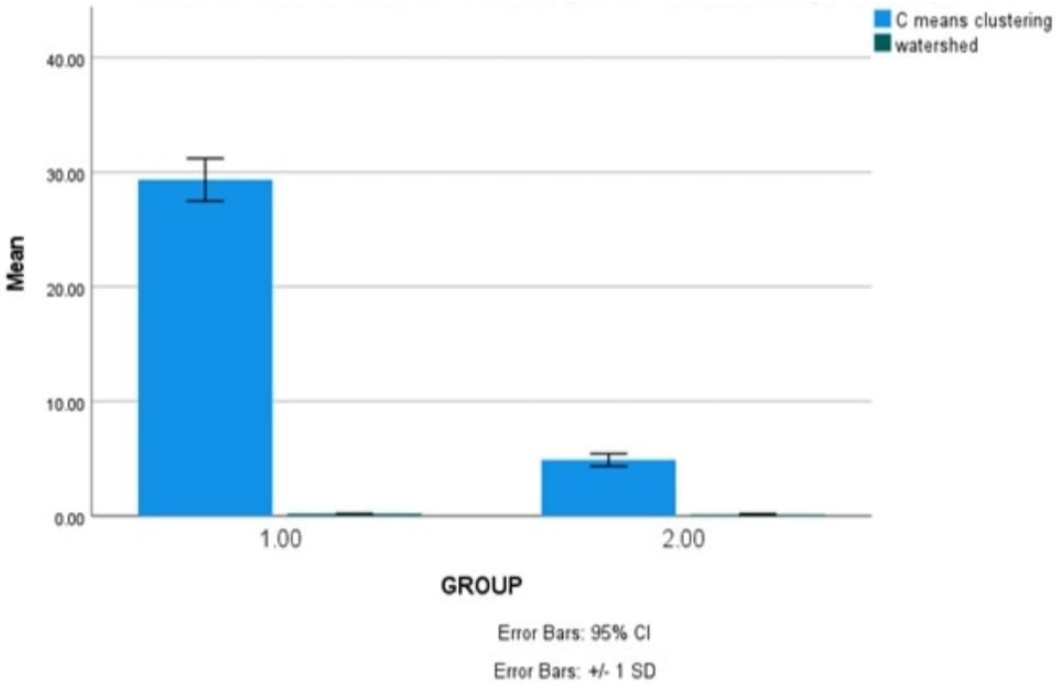
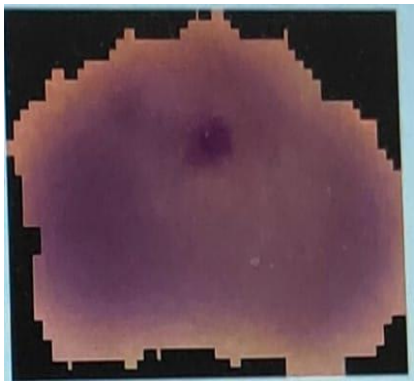
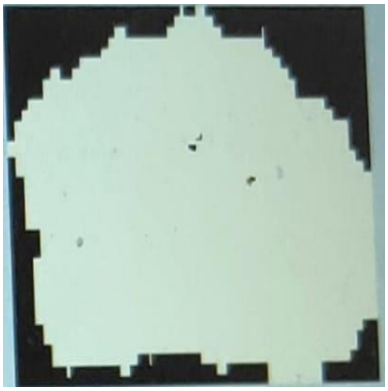


Fig.1. Using the C means clustering algorithm and the Watershed algorithm, we calculated the simple mean of PSNR. The mean PSNR of malaria detection using the Innovative simplified C means clustering algorithm and the Watershed algorithm is represented in a barchart. When compared to the Watershed algorithm, the C-means clustering algorithm appears to yield the most consistent PSNR with the lowest standard deviation. X axis: C means algorithm and thresholding algorithm , Y axis: means PSNR of detection. Mean PSNR of detection ± 1 Standard deviation



Sample image

Fig. 2a. Image obtained using the C Means clustering algorithm. Figure represents a sample image of malaria parasites.



watershed image

Fig. 2b. Image obtained using the Watershed algorithm. This Figure represents a watershed image of malaria parasites.