International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Kapur and Otsu Strategy Based Segmentation and Classification of Plasmodium Species

Idupulapati Divya, Dr. Ramachandra A C
B. Tech, ECE, Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka, India
Professor & Head, ECE, Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka, India

Article Info

Publication Issue
Volume 10, Issue 1
January-February-2023

Page Number
127-133

Article History
Accepted: 01 Jan 2023
Published: 17 Jan 2023

ABSTRACT

Segmentation is one of the important steps for image analysis. Multilevel thresholding image segmentation was more popular in image segmentation. Otsu and Kapur based methods are most popular for multilevel threshold image segmentation. Many authors implemented evolutionary algorithms for the optimal multilevel threshold selection based on the above methods. In this work, an efficient approach for multilevel image segmentation has been proposed and implemented based on Kapur and Otsu strategy. After the segmentation process, classification will be performed on Plasmodium species using Machine Learning algorithm- Random Forest. The types of plasmodium species are P. Falciparum, P. Malariae and P. Vivax which are collected from Kaggle portal. To check the effectiveness of our method/work, image entropy of Kapur and Otsu strategy and classification results are evaluated.

Keywords: - Image Segmentation, Multilevel Thresholding, Kapur and Otsu Strategy, Machine Learning, Classification, Random Forest, Plasmodium Species.

I. INTRODUCTION

In order to assess the various blood cell types, a blood smear is frequently utilised as a follow-up test following abnormal complete blood count (CBC) results. It can be used to support a variety of illnesses that alter blood cell populations in the diagnosis and/or monitoring stages.

Almost everyone who has a CBC was once subjected to a blood smear. It is currently common practise to offer an automated differential due to the advent of more advanced, automated blood cell counting devices. In contrast, a blood smear may be carried out if the outcomes of an automated cell count and/or differential point to the existence of immature, atypical, or abnormal white blood cells (WBCs), red blood cells (RBCs), and/or platelets, or if there is cause to believe that abnormal cells are present.

A blood smear is frequently used to classify and/or detect diseases that affect one or more types of blood cells as well as to keep track of patients receiving therapy for these ailments. The amount and type of blood cells produced, their function, and longevity can all be impacted by a wide range of illnesses, disorders, and nutritional deficiencies. Anemia, myeloproliferative neoplasms, bone marrow issues, and leukaemia are a few examples. Normal, mature, or nearly mature cells are often the sole ones released into the bloodstream, but under certain conditions, the bone marrow may release immature and/or aberrant cells into the bloodstream. When a large number or particular type of aberrant cells are found, it may be a symptom of a disease or condition and lead a doctor to order additional tests.

When a CBC with differential, conducted with an automated blood cell counter, indicates the presence of atypical, aberrant, or immature cells, a blood smear is typically done as a follow-up test. It may also be done if a person exhibits signs and symptoms of a disorder that affects the generation or longevity of blood cells.

When a person is receiving treatment or being observed for a blood cell-related condition, a blood smear may also be requested on a regular basis.

Blood smear examination results are not necessarily diagnostic in and of themselves; more frequently, they point to the potential existence or severity of an underlying illness and the necessity of additional diagnostic tests. The outcomes are taken into account along with the results of the CBC and other laboratory tests, as well as the clinical symptoms and signs of the test subject.

An analysis of the red, white, and platelet look as well as any anomalies that might be visible on the slide are often included in the results of a blood smear.

Red Blood Cells (RBCs)

Normal, adult red blood cells lack a nucleus, unlike the majority of other cells, and are uniformly sized (7-8 μm in diameter). They have a dip in the centre rather than a hole since they are round and flattened like donuts (biconcave). After staining the blood smear, the RBCs have a pink to red colour with a pale core as a result of the haemoglobin they contain. When RBC morphology (appearance) is normal, it is frequently described as normochromic (normal colour) and normocytic (normal size).

Even though not every RBC will be ideal, the presence of disease may be indicated by a sizable number of cells with abnormal form or size.

II. RELATED WORKS

The author of this paper [1] looked into segmentation utilising multi-level thresholding. To identify a threshold level and a new sub-range for the following step, the technique is used repeatedly on sub-ranges derived from the previous step, starting with the extreme pixel values at both ends of the histogram plot, until no appreciable increase in image quality can be made.

The optimal multilevel picture segmentation developed by Kapur utilising the Crow Search Algorithm was employed in this article [2]. With the help of this approach, multi-level thresholding's ideal values are calculated. One of the most popular and basic techniques for picture segmentation is thresholding. In order to divide an image into two segments using bi-level thresholding, or several segments using multi-level thresholding, an image must first be divided into two segments.

This experimental study [3] created and tested the Otsu and Gaussian Otsu thresholding algorithms using a variety of images. In order to establish the threshold value, the performances of the two methodologies' findings were then contrasted. The performance of the Gaussian Otsu's approach is superior to the Otsu's method, according to the results.

This study's [4] author looked into the use of a random forest machine learning classifier to categorise various pieces of data. We provide a detailed examination of a random forests model proposed by Breiman (2004) that is remarkably similar to the original approach in this paper. We specifically demonstrate the consistency and sparsity adaptation of the method by demonstrating that its rate of convergence is independent by the quantity of noise variables and instead depends only on the number of strong features.

The author of this study [5] has researched the numerous datasets that can be collected as well as how to go about doing so. organising the dataset using various categories. I did research on the various species that bring on diseases like malaria.

This article [6] examines the foundations of image processing and its various uses. Improvements in pictorial information for human interpretation and processing of image data for storage, transport, and representation for autonomous machine perception are the two main application areas driving interest in digital image processing techniques. We learn why we refer to this kind of work as image processing in this study. How did image processing begin, and what are the primary study areas? Various books that can teach you about image processing. How may image processing be used in various industries and applications.

III.METHODOLOGY

Four tasks comprise this work:

- Kapur approach for optimal multi-thresholding of RGB pictures to increase entropy
- Otsu strategy-based segmentation of RGB images for increased entropy in plasmodium species detection
- Otsu segmentation approach for plasmodium species classification in thin smear images
- Using the Kapur segmentation approach, plasmodium species are classification in a blood smear image.

Kapur's Entropy-based Segmentation: In the field of image processing, image segmentation is typically used to eliminate the enthusiastic portion of a digital image outline. It is a fundamental advancement in image preparation that aids in isolating an image into non-covering, homogenous zones encasing related objects. Writing about images provides information on the many segmentation strategies that have been presented and implemented by the great majority of analysts.

There are two types of image thresholding techniques: global dimension edge and local dimension limit. While a single edge esteem is given to the entire image when using global dimension thresholding, different edge esteems are designated for each segment of the image when using local dimension thresholding. In this process, a likelihood thickness capacity of the grey dimension histogram is used to determine the maximum incentive with the aid of a parametric or non-parametric methodology. It is confusing and laborious to use image thresholding that depends on parametric approach. The image quality and starting circumstances also had an impact on the system's final outcome. Thus, the majority of experts accept non-parametric approaches to understand the problem of grey and shading image division.

The image is divided into two (bi-level) or more classes using thresholding techniques (RGB). The parallel dimension thresholding involves taking just one edge esteem (t), checking each pixel with an explicit force esteem, and if it is higher, setting the limit esteem (t) as the top of the line and designating the other pixel with a different force esteem as inferior. In multilevel thresholding, the pixels in the image are divided into several portions, with one of the areas being chosen as the most intriguing and having the highest limit. The best edge esteem can be determined fundamentally using two approaches known as parametric and nonparametric. A few characteristics of a likelihood thickness capacity should be evaluated in parametric methodology in order to rank the image class categories.

The computing cost of this methodology is high. A few criteria, such as the error rate, entropy, and other laborious ones are advanced by non-parametric approach to determine the ideal limit values. For paired dimension thresholding, Otsu's and Kapur are the two methods that can be used. Of the two strategies, Otsu entails a shift in the classes; in particular, it observes the greatest change among the classes. For measuring class homogeneity, the Kapur technique, on the other hand, enhances entropy. The limits of these two multilevel threshold approaches are increasing. The goal of this research is to evaluate the efficacy of entropy-based thresholding algorithms and how they combine in a system for identifying cutaneous melanoma lesions. This is achieved by:

- Constructing a novel segmentation method that takes a plasmodium species image and produces a sectioned image with a distinct cell,
- The developed melanoma location technique employs novel entropy-based thresholding,
- combining several thresholding strategies by using legal administrators on the thresholding yield and obtaining a programmed selection of their yields to acquire the greatest result.
- analysing the entropy outcomes and their unique mixtures in the melanoma location that was developed.

Kapur's entropy work was first put forth in 1985 to segment the grey scale image by increasing the entropy of the histogram in order to determine the threshold. The Kapur's entropy at that time will be Jmax = fkapur (Th) = $\sum_{j=1}^{k} H_j^c \dots$ (1)

Most of the time, each entropy is handled independently based on the particular Th esteem. The threshold values are arbitrarily changed by the HS based search till Jmax is met.

Binarization of images is the process of changing a grayscale image to a binary image. The only two potential intensities for pixels in a binary image (I) are nominally 0 and one as in the equation. To distinguish the foreground object from the background, the image must be binarized. According to the threshold calculation approach, the binarization techniques can be divided into many groups.

$I : (x, y) \to \{0,1\}$

Thresholding: By giving each pixel an intensity value, thresholding is utilised to separate an object from its background and determine if the pixel is an object or a background. Global thresholding employs a suitable threshold (T)

$$g(x,y) = \begin{cases} 1, if\ f(x,y) > T \\ 0, if\ f(x,y) \leq T \end{cases}$$

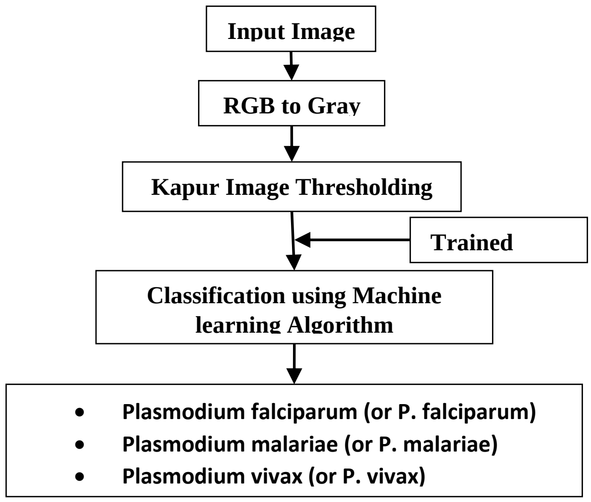

Fig 1: Block Diagram of Proposed Method Kapur Segmentation

An efficient method of image segmentation is developed using Kapur's entropy. In order to evaluate the effectiveness of the suggested strategy, several images, including natural images, satellite photos, and magnetic resonance (MR) images, are selected for experiments. The number of bits required to encode visual data is measured by discrete entropy, which is employed in image processing. The entropy value will increase as the image becomes more detailed.

Thresholding at multiple levels

All image thresholding methods fall into one of two basic categories: bi-level thresholding methods or multi-level thresholding methods. Bi-level thresholding approaches divide an image into two classes—foreground and background—using a single threshold value, but they perform less well when the scene is complex and contains a wide range of objects. In order to segment images, multilevel thresholding techniques are widely used. The threshold values in this study are determined using Kapur's entropy, a well-known multilevel thresholding technique. The method is succinctly described in the following subsections. Additionally, these thresholding algorithms are applied three times to obtain the appropriate threshold values for each of the three colours since the RGB image is made up of the three primary colours red, green, and blue.

The Kapur's Entropy

In Kapur's method, which is also an unsupervised automatic thresholding approach, the optimal thresholds are chosen based on the entropy of segmented classes. Assuming that many thresholds represent the thresholds that, when combined, divided the image into several classes.

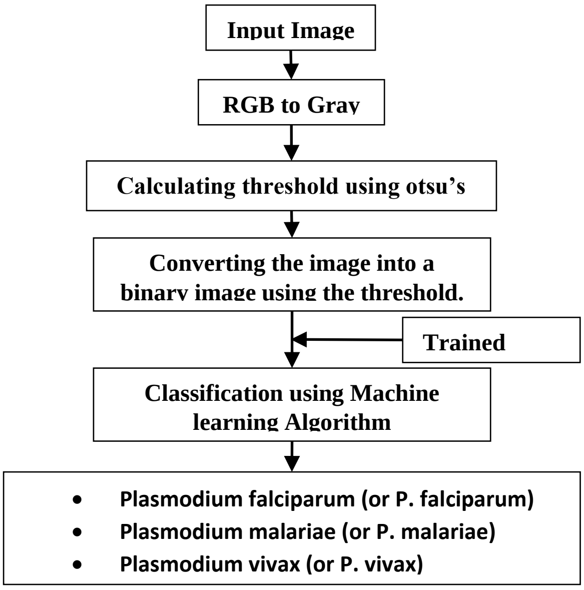

Fig 2: Block Diagram of Proposed Method OTSU's Segmentation

Otsu's approach, named after Nobuyuki Otsu, is employed in computer vision and image processing to carry out automatic image thresholding. [1] The algorithm returns a single intensity threshold in its most basic form, dividing pixels into the foreground and background classes. This threshold is established by maximising inter-class variation or, alternatively, minimising intra-class intensity variance.

A global adaptive binarization threshold image segmentation algorithm is the OTSU method (OTSU). The threshold selection rule for this approach uses the highest interclass variance between the background and the target image. Using an iterative process, Otsu's thresholding approach determines a measure of spread for the pixel levels on either side of the threshold, or the pixels that are either in the foreground or background.

The most effective technique for global thresholding is Otsu's thresholding. To transform a grayscale image into a binary image, it automatically conducts histogram shape-based image thresholding. The approach, which is based on image variance, determines the best threshold by maximising the variation across classes. The ideal threshold value (t*) satisfies these conditions:

$$\sigma_B^2(t^*) = max_{0<t<L-1}\sigma_B^2(t)$$

Where σ_B between classes variances, L intensity levels [0, ---, L-1]

Random Forest: The supervised learning method Random Forest is utilised for both classification and regression. But it is primarily employed for classification issues. As is common knowledge, a forest is made up of trees, and a forest with more trees will be more sturdy. Similar to this, the random forest algorithm builds decision trees on data samples, obtains predictions from each one, and then uses voting to determine the optimal option. Because it averages the results, the ensemble method—which is superior to a single decision tree—reduces over-fitting. The Random Forest Algorithm's ability to handle data sets with both continuous variables, as in regression, and categorical variables, as in classification, is one of its most crucial qualities. In terms of classification issues, it delivers superior outcomes. A feature descriptor called HOG, or Histogram of Oriented Gradients, is frequently employed to extract features from image data. It is commonly used for object detection in computer vision tasks. The HOG description emphasises an object's structure or shape. We only determine if a pixel is an edge in the case of edge features. HOG is also able to offer the edge direction. To do this, the gradient and orientation (or magnitude and direction, depending on your preference) of the edges are extracted.

Using one of the machine learning algorithms called Random Forest Classifier and the extraction of HOG (Histogram of Oriented Gradients) features, we will classify whether the input image is a plasmodium species in this project. This model uses the Kaggle plasmodium species Dataset for testing and training.

Working of Random Forest Algorithm:

We can understand the working of Random Forest algorithm with the help of following steps –

- Step 1: Choose random samples at random from the provided dataset.
- The second step of this technique is to create a decision tree for each sample. The forecast outcome from each decision tree will then be obtained.
- Step 3: Voting will be conducted for each predicated outcome.
- Step 4: Finally, choose the prediction result that received the most votes as the final prediction result.

IV. RESULTS AND DISCUSSION

The below figures are the experimental outputs of the proposed method

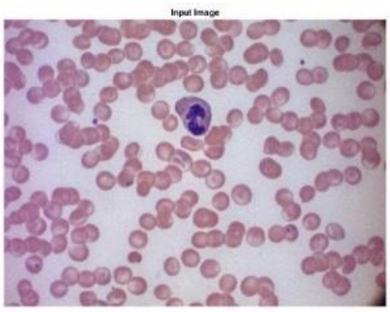

Fig 3: Input Image

First, we will select a test image as input image.

Fig 4: Segmentation using Kapur Strategy

Here we apply kapur strategy for segmentation of plasmodium species and finds entropy values for the segmented image.

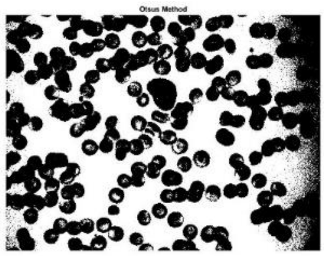

Fig 5: Segmentation using Otsu Strategy

Then we apply Otsu's Method for segmenting the input image after that calculates entropy values.

```
Image Entropy of Kapur Strategy:
    0.0561

Image Entropy of Otsu Strategy:
    0.9994

Classification output of Kapur Strategy:
P.Falciparum
Classification output of Otsu Strategy:
P.Falciparum
```

Fig 6: Entropy and output classification of Plasmodium species

Here, the output is a classification of the plasmodium species type and the derived entropy values. With the help of hog characteristics collected from segmented images, the classification is provided by a random forest classifier.

V. CONCLUSION

Here, we've done image segmentation using thresholding approaches like the Kapur and Otsu ones. These studies are employed in the classification process, which divides the three different parasites into plasmodium falciparum, malariae, and vivax. By keeping an eye on the outcomes of an experimental session, we may assess the models and methodologies.

VI. REFERENCES

[1]. S. Arora, J. Acharya, A. Verma, Prasanta K. Panigrahi, “Multilevel thresholding for image segmentation through a fast statistical recursive algorithm”, Received 12 February 2006; received in revised form 1 July 2007 Available online 22 September 2007.

[2]. Pankaj Upadhyay, Jitender Kumar Chhabra, “Kapur’s entropy based optimal multilevel image segmentation using Crow Search Algorithm”, Applied Soft Computing, Volume 97, Part B, 2020, 105522, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2019.105522

[3]. Yousefi, Jamileh. (2015). “Image Binarization using Otsu Thresholding Algorithm”. 10.13140/RG.2.1.4758.9284.

[4]. Gerard Biau, “Analysis of a Random Forests Model”, Journal of Machine Learning Research 13 (2012) 1063-1095

[5]. Saif Uddin, “Malaria Parasite Image (Different Malaria Species)”, Kaggle, Sun Jun 07 2020, https://www.kaggle.com/saife245/malaria-parasite-image-malaria-species

[6]. Rafael C. Gonzalez, Richard E. Woods, “Digital Image Processing”, Person International Edition prepared by Pearson Education

Cite this article as :

Idupulapati Divya, Dr. Ramachandra A C, "Kapur and Otsu Strategy Based Segmentation and Classification of Plasmodium Species", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 1, pp. 127-133, January-February 2023.
Journal URL : https://ijsrst.com/IJSRST231018