International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 11
ISSN: 2454-4248
350 – 354

A Study of Dengue Infection Segmentation, Feature Extraction and Classification

Ragini Deshmukh
ME Student
Computer Engineering Department
SIE, Vadodara
raginideshmukh555@gmail.com

Sheshang D. Degadwala
Head of Department
Computer Engineering Department
SIE, Vadodara
sheshang13@gmail.com

Arpana D. Mahajan
Assistant Professor
Computer Engineering Department
SIE, Vadodara
arpana.mahajan@yahoo.com

Abstract: Aedesaegypti mosquito spared the dengue viral illnesses. The world's greatest developing outbreak is dengue fever. Day by day the rate of dengue has become significantly around the globe increases. Dengue infections are of three forms: Dengue fever additionally perceived as "break bone" fever, Dengue Haemorrhagic Fever (DHF), Dengue Shock Syndrome (DSS) which are life debilitating. Doctors need to capture approximately 20 to 50 pictures of white blood cell from different angle to identify the disease. The platelet count is estimated using various segmentation techniques and morphological operations with the help of the platelets count dengue fever infection is detected. A technique used for segmentation are mainly thresholding based that is not segment exact part of defected platelet. But, the result was not so efficient in providing the spatial detail information of the actual disease part. So here we are going to use Fuzzy based algorithm to segment WBC Platelets. There are different feature extraction methods are apply platelet are size, shape and area. But it was not giving the exact results. So here we are going to use Haarlick Features for WBC platelets. And any machine learning method SVM, ANN, Decision Tree will be used for the classification of dengue infection types.

Keywords: *Dengue Fever, Segmentation, feature Extraction, Classification and Comparative Analysis.*

*****

I. INTRODUCTION

The Blood Cell has been observed in many patient. The Exudate Dengue Fever is mainly divided into three types. Doctors need to capture approximately 20 to 50 pictures of WBC from different angle to identify the types and its Counts. still, the result was not so efficient in providing the spatial detail information of the actual disease part. The research further here, uses Color and Clustering based methods have been proposed and modified to segment the WBC from blood cell image and provide the spatial detail of segmented part. Extraction of WBC features and Classification of types of Dengue is done using the Machine Learning Tool.

A blood cell disorder is a condition in which there's a problem with your red blood cells, white blood cells, or the smaller, circulating cells called platelets, which are critical for clot formation. All three cell types form in the bone marrow which is the soft tissue inside your bones.

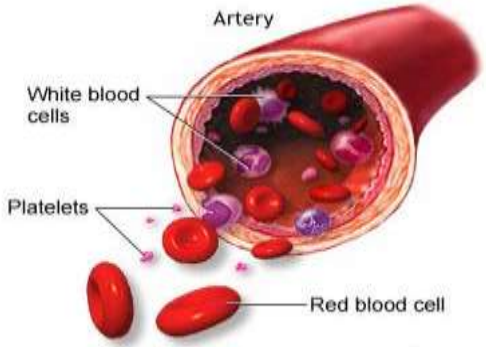

Figure 1.1 White Blood Cell

Classical dengue fever (DF) is a viral illness transmitted through the bite of an infected mosquito, usually Aedes aegypti or Aedes albopictus (Halstead, 1988). This illness is endemic in tropical regions and affects between 50-100 million people worldwide annually (Gibbons & Vaughn, 2002).Dengue illness can range from a nonspecific febrile illness, as in DF, to a more severe illness with bleeding tendency, thrombocytopenia, and plasma leakage (dengue hemorrhagic (DHF), Dengue Shock Syndrome (DSS)).

Table 1.1 Patients and clinical classification of dengue

Table 1 - Patients and the clinical classification of dengue

Type	Age < 15 years n (%)	Age ≥ 15 years n (%)	Total
Classic dengue fever	4 (2.60)	16 (10.39)	20
Severe dengue fever	18 (11.69)	26 (16.88)	44
Dengue hemorrhagic fever	44 (28.57)	46 (29.87)	90
Total	66 (42.86)	88 (57.14)	154

A p-value = 0.06 (chi-square test) shows that there is no significant difference in the proportion of clinical forms of dengue fever between under 15-year-old patients and older patients

After studying and analysing various papers, it is found that many work have been done on Segmentation of WBC. But till now the no methods exists for accurate dengue type classification using Machine Learning. So, it is necessary to avoid such kind of limitations by using the method that can work within such scenario where, it is possible to classify the dengue fever type for any blood cell image to improve the accuracy of the system with all kind of environments.

 ISSN: 2454-4248
350 – 354

II. LITERATURE SURVEY

This Review is divided in two sections, first section discusses about the theoretical background of major area of this work that is WBC Segmentation, Feature Extraction and Classification techniques. Second sections gives brief summary of various previous discussions on related work annotated in various research papers.

[I] Segmentation Methods

A. K-means Clustering

K-means clustering is an iterative technique that is used to segregate the image into clusters. It focuses on segregating and grouping the pixels into cluster which has a nearest mean intensity value. The steps of the algorithm is shown below:

Step 1: Read the image
Step 2: Convert the image from RGB color space to L*a*b color space
Step 3: Classify the colors in a*b* space using k-means clustering.
Step 4: Label every pixel in the image using result from k-means.
Step 5: Create images that separate the HE image by color.
Step 6: Segment the nuclei into a separate image.

B. Fuzzy C-means (FCM)

The individual's feathery c-means (FCM) grouping calculation might need been regardless familiar at Dunn [17] Also following the way extended Toward Bedeck. The individuals count might a chance to be a iterative grouping framework that produces a Perfect section at minimizing the weighted inside amassed aggregate to squared slip goal ability.

$$J_m = \sum_{i=0}^{N} \sum_{j-1}^{e} u_{ji}^{m}\, d^2\,(x_i\,, v_j\,)$$

Those spot X = {x1, x2… xn} ⊆ Rm might make the data set in the m dimensional vector space, n is the individuals measure for majority of the data items,. C may be the measure for aggregations with 2 < c's < N, Uij might a chance to be the level of enlistment starting with asserting Xi in the Jth cluster, m will a chance to be the individuals weighting kind When every feathery membership, Vj is the individuals model of the concentrate over bunch, may be an detachment measure those center from claiming thing Xi Moreover one assembly centering Vj. A address of the fill in Jm Might aggravate procured through An iterative process, which will be passed on similarly takes then afterward.

1. Situated qualities for ,m , Furthermore €.
2. Instate the fluffy segment grid U(0).
3. Situated those circle counter b = 0.
4. Ascertain the bunch centres vj(b) for U(b).

$$v_j^{(b)} = \frac{\sum_{i=1}^{N}(u_{ji}^{(b)})x_i}{\sum_{i=1}^{N}(u_{ji}^{(b)})^m}$$

5. Compute the membership matrix U(b+1)

$$u_{ji}^{(b+1)} = \frac{1}{\sum_{k=1}^{c}(\frac{d_{ji}}{d_{ki}})^{2/m-1}}$$

6. If max {U(b) – U(b+1)} < € that point stop, generally set b = b+1 Also try should step 4.

C. LAB Color Segmentation

The Lab color space has three channels- L is the lightness channel, A and B are the two color channels. The L channel has values ranging from 0 up to 100, which correspond to different shades from black and white. In Figure 2.3 the A channel has values ranging from -128 to +127 and gives red to green ratio and the B channel has values ranging from -128 to +127 and gives yellow to blue ratio[7]. Thus high value in A or B channel represent a color having more red or yellow and a low value represent a color having more green or blue.

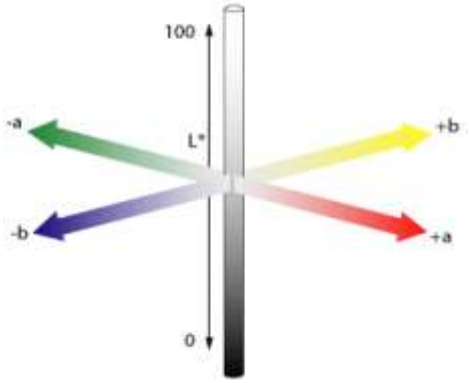

Figure 2.1 Illustration of the LAB color space

D. CMYK Color Segmentation

CMYK model uses the secondary color of light cyan, magenta and yellow. Some printers and devices use secondary colors instead of the primary colors. The conversion from RGB to CMY can be performed as follows:

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Mixing of all the three secondary colors in equal amount produce the black color, but for printing purpose it does not produce the true black. So, the additional black is added with these secondary color gives rise to the CMYK color model.

[II] Feature Extraction Methods

A. Shape

[1] Area
Returns a scalar that specifies the actual number of pixels in the region.
[2] Perimeter

 ISSN: 2454-4248
350 – 354

Returns a scalar that specifies the distance around the boundary of the region. It computes the perimeter by calculating the distance between each adjoining pair of pixels around the border of the region.

[3] Major Axes

Returns a scalar that specifies the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

[4] Minor Axes

Returns a scalar that specifies the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

B. Texture

[1] GLCM

These features capture information about the patterns that emerge in texture of an image. The features are calculated by construction of a Gray Level Co-occurrence Matrix (GLCM).The GLCM calculates how often a pixel with gray-level (grayscale intensity) value i occurs either horizontally, vertically, or diagonally to adjacent pixels with the value j. GLCM direction of analysis:

- Horizontal (0°)
- Vertical (90°)
- Diagonal
- Bottom left to top right (-45°)
- Top left to bottom right (-135°)

Denoted as P_0 , P_{45} , P_{90} & P_{135} respectively.

Ex : P_0 (i,j) where i and j are gray level values (tones) in the image.

Example of Directional Analysis P_0 , P_{45} , P_{90} & P_{135} .

[2] Haarlick

There are three kinds of visual cues people naturally look for in an image: spectral (average tonal variation in various bands of visible wavelengths), contextual (macro data surveyed from surrounding data), and textural. Textural information, or the spatial distribution of tonal variation within a band, is one of the most important characteristics used in identifying objects or regions of interest in an image.Haralick, Shanmugam, and Dinstein introduced a set of 13 texture features calculated from an image's gray-level co-occurrence matrix (GLCM).These Haralick features, which are still widely used today for a range of applications, allow quantification of a texture.

The features are given as:

1. Angular Second Moment
2. Contrast
3. Correlation
4. Sum of Squares: Variance
5. Inverse Difference Moment
6. Sum Average
7. Sum Variance
8. Sum Entropy
9. Entropy
10. Difference Variance
11. Difference Entropy
12. Information Measures of Correlation
13. Maximal Correlation Coefficient

[III] Classification Methods

[1] Support Vector Machine

SVM is used to train the data sets using Support Vectors. These Support Vectors are then used to create Hyper planes. Hyper plane then classifies the objects into two classes. One is the positive and another one is negative.

[2] Neural Network

A neural system is an information transforming framework comprising of a vast number of simple, exceptionally interconnectedness transforming components for a structural engineering propelled by those structure of the cerebral cortex part of the cerebrum. Hence, neural networks are often fit about finishing things which people alternately animals do well At which routine workstations frequently all the would poorly. Neural networks bring developed in the previous couple quite some time likewise a territory about surprising chance for research, improvement Furthermore requisition should an assortment for true issues. Indeed, neural networks show aspects Furthermore abilities not given by whatever viable innovation organization. Cases incorporate perusing Japanese Kanji characters Also human handwriting, perusing typewritten text, compensating to arrangement errors done robots, translating thick, as "noisy" signs (e. G. Electrocardiograms), demonstrating intricate frameworks that can't be modelled mathematically, What's more foreseeing if recommended credits will be great alternately come up short.

[3] Decision Tree

Decision tree algorithms is a method to learn decision by means of machine learning techniques to learn decision procedures. The model in which includes predictor variable and response variable is a supervised learning. The tree structure are root node, leaf node and branch node, as shown Figure. Branch node is separated from the root node to the conditions until leaf node which is the last node. The condition of the root node to the leaf node is a pattern of each predictor variable as shown in (2), Information Gains represents the level of a node. Which an Entropy value is the separation of the branch node. Calculation as show in (3). The most value of Information Gain is the root node of the model.

III. COMPARATIVE ANALYSIS

Table 1.2 Segmentation Techniques

Techniques	Advantages	Disadvantages
Fuzzy Clustering FCM	+Enhance clustering performance. +Image detail preservation.	-Apriori specification of no. of clusters. -increase no. of iteration. -sensitive to noise.

 ISSN: 2454-4248
350 – 354

Clustering K-means	+Works well for noise free images.	-Works on fixed distance. -Objective function is defined previously.
Color LAB, CMYK, HSV	+It gives optimized result for features extractions associated with image pixels.	-Problems connected cells closely related to overlapping and skeletonizing.
Hybrid CMYK-LAB-K-means	+Works with all Light Conditions. +High Accuracy +Less Complex. +Low Memory and time Consumption.	+Post-Processing is Need after segmentation.

Table 1.2 Feature Extraction Techniques

Techniques	**Advantages**	**Disadvantages**
Shape		
Area, Perimeter, Major and Manor Axis	Easy to implement, Less Complex, Less Time Consuming	Works with Binary Image only, feature value change when change image dimension
Texture		
GLCM Feature	Computation Time is Low, Low memory Consumption	Works with Gray scale images, feature vector is low so classification accuracy will less.
Haralick Feature	Computational accuracy of feature vectors is high, Classification accuracy is high	Due to 13 features the computation of feature vectors is complex and time consuming.

Table 1.2 Classification Techniques

Techniques	**Advantages**	**Disadvantages**
Decision Tree	-It reduces overfitting and is therefore more accurate. -Easy to Implement -works with all types of data. -Multi classification Support.	It may not work if the dependent variables considered in the model are linearly related. Therefore one has to remove correlated variable by some other technique.
ANN	+High degree of non-linearity possible.	-Hard to tune parameters. -Takes time to build model.
SVM	+High accuracy +Easy to generate rules. +Easy to understand.	-Hard to interpret -It takes more time to predict the new instance.

IV. CONCLUSION

Segmentation is considered an important step in the automatic diagnosis of different computer systems. It was found that several methods in the literature have shown promising results. We propose color and fuzzy cluster based segmentation method and extract the texture features (i.e. Contrast, Correlation, Energy, Homogeneity etc.) from segmented image and the relevant features are used in supervised classifier to classify Dengue Fever.

REFERENCES

[1] J. Poornima and K. Krishnaveni "Detection of Dengue Fever with Platelets Count using Image Processing Techniques" Indian Journals of Science and Technology, Vol 9(19), May 2016

[2] Zhi Liu 1, Jing Liu "Segmentation of White Blood Cells through Nucleus Mark Watershed Operations and Mean Shift Clustering" Sensors 2015, 15, 22561-22586; doi:10.3390/s150922561

[3] Sarach Tantikitti, Sompong Tumswadi, Wichian Premchaiswadi "Image Processing for Detection of Dengue Virus based on WBC Classification and Decision Tree" Thirteenth International Conference on ICT and Knowledge Engineering 2015

[4] Khan, S., Khan, A., Saleh Khattak, F. and Naseem, A. "An Accurate and Cost Effective Approach to Blood Cell Count" International Journal of Computer Applications, 50(1), pp.18-24. 2012

[5] Viswanathan, P. "Fuzzy C Means Detection of Leukemia Based on Morphological Contour Segmentation" Procedia Computer Science, 58, pp.84-90 2015

[6] Narjes Ghane, Alireza Vard, Ardeshir Talebi, and Pardis Nematollahy "Segmentation of White Blood Cells From Microscopic Images Using a Novel Combination of K-

ISSN: 2454-4248
350 – 354

Means Clustering and Modified Watershed Algorithm" J Med Signals Sens. 2017 Apr-Jun; 7(2): 92–101

[7] Safuan, S. N., Tomari, M. R., & Zakaria, W. N. (2017). White Blood Cell (WBC) Counting Analysis in Blood Smear Images Using Various Color Segmentation Methods. Measurement. doi:10.1016/j.measurement.2017.11.002

[8] ogado, L. H., Veras, R. D., Andrade, A. R., Silva, R. R., Araujo, F. H., & Medeiros, F. N. (2016). Unsupervised Leukemia Cells Segmentation Based on Multi-space Color Channels. 2016 IEEE International Symposium on Multimedia (ISM). doi:10.1109/ism.2016.0103

[9] Robert M. Haralick, K. Shanmugam, And Its'hak Dinstein "Textural Features for Image Classification" IEEE transactions on systems, man, and cybernetics, \o smc(-3, no. 6, November 1973

[10] Gautam, A., & Bhadauria, H."White blood nucleus extraction using K-Mean clustering and mathematical morphing" 5th International Conference - Confluence The Next Generation Information Technology Summit (Confluence). doi:10.1109/confluence.2014.694922

[11] Savkare, S. S., & Narote, S. P. "Blood cell segmentation from microscopic blood images"International Conference on Information Processing (ICIP). doi:10.1109/infop.2015.7489435

[12] Mondal, P. K. "Segmentation of White Blood Cells Using Fuzzy C Means Segmentation Algorithm" IOSR Journal of Computer Engineering, 16(3), 01-05. doi:10.9790/0661-16390105 2014