Detection of Dengue Fever by Using K-Means Clustering With SVM Classifier

A. PRABHAVATHI[1], B. CHANDRA LEKHA[2], AAMANI. N[3], A. NAVYA SANTOSHI[4]
[1, 2, 3, 4] *Student, Department of ECE, Vasireddy Venkatadri Institute of Technology, Guntur, India*

Abstract- ***Dengue is a major health problem in tropical and Asia-Pacific regions. Dengue is a mosquito-borne infection that can lead to a severe flu-like illness. It is caused by four different viruses and spread by Aides mosquitoes. According to World Health Organization (WHO) every year nearly 400 million people are affected by Dengue Fever. This project uses blood smear images with white blood cells classification. Here based on White Blood Cells (WBC) classification it is going to detect whether a person is infected with dengue or not. The proposed method is implemented with the help of K-Means clustering with Support Vector Machine (SVM) classifier. The proposed K-Means clustering with SVM classifier gives better results.***

Indexed Terms- blood smear images, K-Means clustering, SVM classifier.

I. INTRODUCTION

Dengue fever is a mosquito-borne tropical disease caused by the dengue virus. Symptoms typically begin three to fourteen days after infection. These may include high fever, headache, vomiting, muscle and joint pains, and a characteristic skin rash. Recovery generally takes two to seven days. In a small proportion of cases, the disease develops into severe dengue, also known as dengue hemorrhagic fever, resulting in bleeding, low levels of blood platelets and blood plasma leakage, or into dengue shock syndrome, where dangerously low blood pressure occurs [1].

Dengue is spread by several species of female mosquitoes of the Aedes type, principally Aaegypti. The virus has five types; infection with one type usually gives lifelong immunity to that type, but only short-term immunity to the others. Subsequent infection with a different type increases the risk of severe complications. A number of tests are available to confirm the diagnosis including detecting antibodies to the virus or it's RNA. Dengue has become a global problem since the Second World War and is common in more than 120 countries, mainly in Asia and South America. About 390 million people are infected a year [2] and approximately 40,000 die.

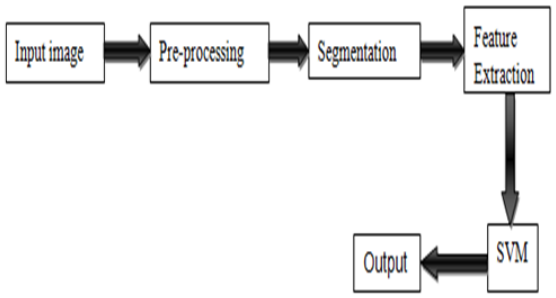
The existing method is based on Artificial Neural Networks (ANNs) and having a limited accuracy, precision, sensitivity and specificity of 80% in its performance evaluation metrics.

II. PROPOSED METHOD

With the technological advances in medical field, the need for faster and more accurate analysis tools becomes essential for better patient's diagnosis. In this work, the image recognition problem of white blood cells (WBC) is investigated. Five types of white blood cells are classified using a feed forward back propagation neural network [4]. After segmentation of blood cells that are obtained from microscopic images, the most 16 significant features of these cells are fed as inputs to the neural network. Half of the 100 of the WBC sub-images that are found after segmentation are used to train the neural network, while the other half is used for test. The results found are promising with classification accuracy being 96%.

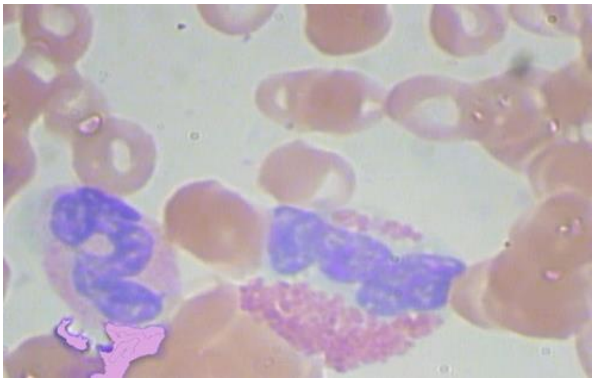
So here for the detection purpose we are using of K-Means Clustering with Supporting Vector machine (SVM).

III. BLOCK DIAGRAM



IV. INPUT IMAGE

For the detection of dengue fever here we use blood smear images [3] that are taken under a digital microscope with 400X magnification specifications by means of image pre-processing techniques such as image acquisition, pre-processing, image segmentation, edge detection feature extraction and white blood cells classification.



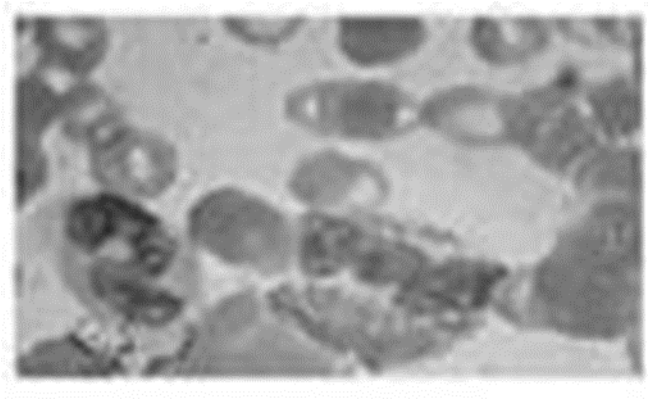
Input image

V. PRE-PROCESSING

Pre-processing helps in improvement of the image data by suppressing unwanted distortions.
In the microscopic images which are given as input image consists of Gaussian noise and shot noise. Gaussian low pass filter is used to remove the noise in the image and smooth the image for further process which is segmentation.
Gaussian function is represented as follows:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

Then the output from the pre-processing block is



Pre-processed image

VI. SEGMENTATION

The division of an image into meaningful structures is called as image segmentation. It is often an essential step in image analysis, object representation, visualization, and many other image processing tasks.

There are various techniques for image segmentation but in this project K-means clustering is used for image segmentation [5].

VII. K-MEANS CLUSTERING

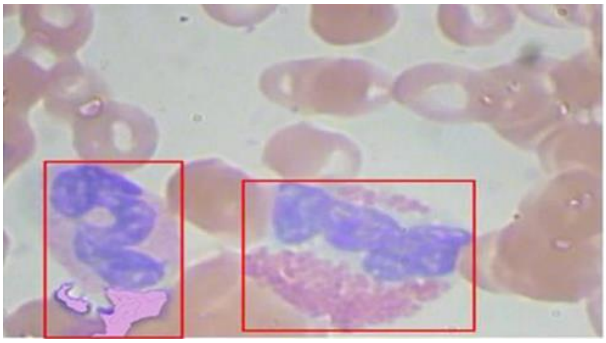
Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different [9].

We use it here to denote techniques that are primarily used in exploratory data analysis of high-dimensional measurement patterns. This K-Means clustering is done through K-Means algorithm.

VIII. K-MEANS ALGORITHM

K-means algorithm is an iterative algorithm that tries to partition the dataset into *K* pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the inter-cluster data points as similar as possible while also keeping the clusters as different (far) as possible.

The way k-means algorithm works is as follows:
Step 1: Specify the number of clusters (k).
Step 2: Allocate objects to clusters.
Step 3: Compute cluster mean.
Step 4: Allocate each observation to the closest cluster centre.
Step 5: Repeat steps 3 and 4 until the solution converges.
The output from the segmented block is as follows:



Segmented image

IX. FEATURE EXTRACTION

Feature extraction is done based on intensity, shape and texture.

Texture is a measure of the variation of the intensity of a surface, quantifying properties such as smoothness, coarseness and regularity [6].

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level occurrence matrix (GLCM) [8], also known as the gray-level spatial dependence matrix.

X. SVM

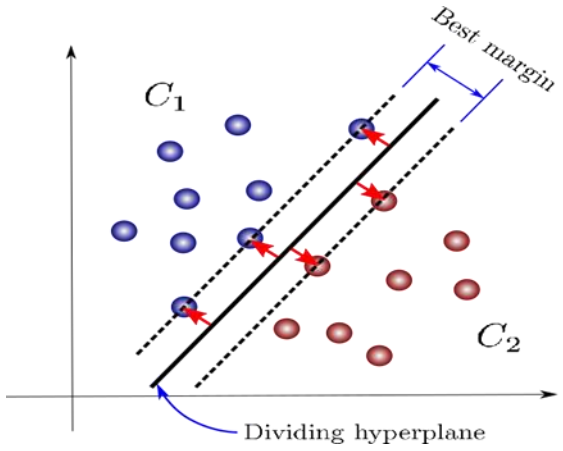
Support Vector Machine is a supervised machine learning algorithm which can be used for both classification and regression challenges.

The objective of the support vector machine algorithm is to find a hyper plane in an N-dimensional space (N- the number of features) that distinctly classifies the data points [10].

The chosen plane must have a maximum margin, i.e., the maximum distance between data points of both classes.

Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence [7].

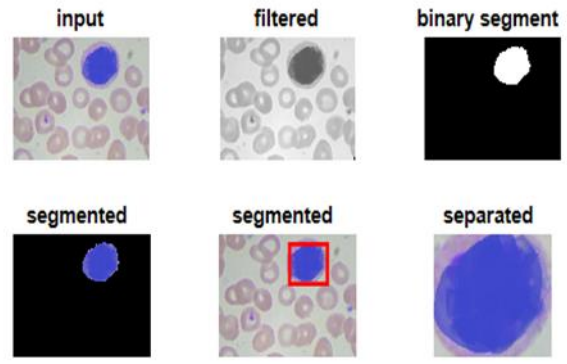
The hyper plane can be computed as follows,
$h(p)= y^T p + y_0 = 0.$

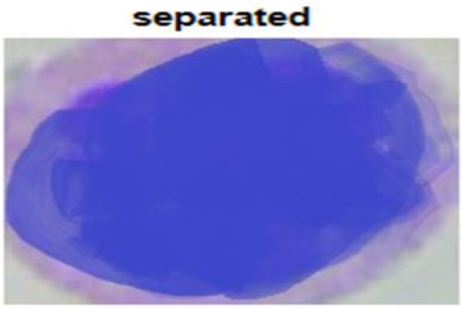


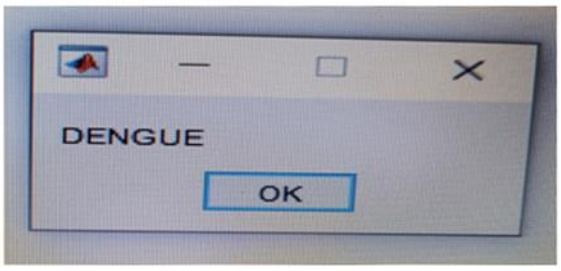
XI. REPRESENTATION OF SUPPORT VECTORS

The feature extracted image is then processed through SVM and the WBC gets separated and that separated image is used in order to define when the person is affected with dengue or not and the results are as follows:

XII. RESULTS







CONCLUSION

In this paper we detect the dengue fever by using K-Means clustering in which segmentation of images is done with an accuracy of 97% and SVM classifier classifies the images with an accuracy of 90-95%. Mainly it is easy to identify as we are using WBC cells for the detection purpose. The above two techniques defines whether the person is infected with dengue or not with an accuracy of 90-95%.

REFERENCES

[1] Dengue Fever. 15 Jan 2016. Available from:http://www.nhp.gov.in/disease/musculo-skeletal-bone-joints-/dengue-fever

[2] Dengue Fever. 15 Jan 2016. Available from:https://en.wikipedia.org/wiki/Dengue_ fever#Cause

[3] Blood Smear preperation procedure. 12 Jan 2015. Available from: https://en.wikipedia.org/wiki/Blood_ film#Preparation

[4] T. S. Mohammed and Nidal Ibrahim al-Tataie, “Artificial Neural Network as a Decision- Makers for Stereo Matching”, GSTF- International Journal on Computing Vol. 1, No. 3, pp (89 – 94), August 2011.

[5] M. Z. Othman,., and Ali, Alaa, B.," Segmentation and Feature Extraction of Lymphocytes WBC using Microscopic Images", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181,Vol. 3 Issue 12, December-2014.

[6] A. Gebejes, A. and Huertas, R.," Texture Characterization Based on Grey- Level Co-occurrence Matrix", Granada: Conference of Informatics and Management Sciences, 2013.

[7] Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Systems, Man, and Cybernetics. 1979; SMC- 9(1):62–6.

[8] A. Gebejes, A. and Huertas, R.," Texture Characterization Based on Grey-Level Co-occurrence Matrix", Granada: Conference of Informatics andManagement Sciences, 2013.

[9] K-Means clustering with svm classifier available:https://ieeexplore.ieee.org/document/8528782

[10] A. Ben-Hur, C.S. Ong, S. Sonnenburg, B. Schölkopf, G. Rätsch, "Support vector machines and kernels for computational biology", PLoS computational biology, vol. 4, no. 10, pp. e1000173, 2008.