DENGUE FEVER PREDICTION USING MACHINE LEARNING ALGORITHM

Miss.M.Jothilakshmi[1] Mrs.P.Jasmine Lois Ebenezar[2] and Mrs.E.Julieruth[3]

Department of Computer Applications, Sarah Tucker College, Thirunelveli-7.

ABSTRACT

Dengue fever is a major problem in many developing countries, including India. For dengue patient monitoring, platelet count is vital to ensure early treatment in order to prevent disease complications. In primary health care centers platelet counting is typically performed manually, which is labor intensive and requires an experienced laboratory technician. Another method used, is the Advia hematology analyzer, which is very expensive, not affordable for rural and remote areas. To address present day challenges, developed an automated approach for the detection of platelet, along with the symptoms helps in assisting the detection of dengue. The technology is based on microscopic images derived from blood smears obtained using a digitalized camera attached to a microscope. Image processing and segmentation techniques are applied to estimate the platelet count from these blood slides. To further improve the accuracy of the results, an analysis of symptoms present in the patient is used in conjunction with the platelet analysis. Proposed vector based method for screening the samples. The classifier performance is evaluated with the sensitivity and specificity values. The results of platelet counts obtained from other platelet counting machines and manually are compared. The reported tool helps in the modernization of pathology laboratory into digital, as well it speedup the mass screening.

1. INTRODUCTION

1.1 General Background

Dengue is the mosquito-borne viral disease caused by female mosquitoes called Aedes Aegypti. These are also main vectors of yellow fever, chikungunya and zika viruses. Aedes Aegypti is a type of flaviviridea family they are four distinct but closely related, serotypes of the virus that cause dengue (DENV-1, DENV-2, DENV-3 and DENV-4). The dengue virus is transmitted to persons through the bites of infected female mosquitoes mainly Aedes Aegypti mosquito. The genus of aedes also acts as a vector but their contribution is not much as Aedes Aegypti. Dengue should be suspected when a high fever (40^0C/104^0F) is accompanied and have the symptoms of severe headache, pain behind the eyes, muscle and joint pain, nausea, vomiting, swollen glands and rash. Dengue may be converted into severe dengue that means dengue hemmorgage fever is called critical phase normally about 3-7 days after the illness severe abdominal pain persistent vomiting, rapid breathing, bleeding gums, fatigue, blood in vomit are the warning signs of dengue hemmorgage fever.

Dengue affected 141 countries all over the world. Every year 390 million persons were affected by dengue illness all over the world. In worldwide 25,000 deaths were recorded per annual. In india 70% of peoples affected by dengue. Past ten years the number of dengue cases increased. The person affected by dengue can be identified through platelet count.

Platelets are tiny blood cell have no cell nucleus, they fragment of cytoplasm that are derived from megakaryocytes of the bone marrow. The platelets play a main role in blood clotting. The normal range of platelet count is 150000-350000 per micro litre of blood.

The count is below the level, person have illness. The platelets are also known as thrombocytopenia. It contains granules that can secrete other proteins required for creating a firm plug to seal blood vessel breaks, also contains proteins similar to muscle proteins that allow them to change shape when they become sticky. In dengue the platelet count goes down to 20000 or even lower. Low platelet count does not lead to dengue additionally the symptoms along with count will be proceed the person have dengue.

Platelets are count in the laboratory by the expertise otherwise using a special type of chamber, designed for the counting of blood cells in the specimen, known as Neubauer's chamber or Hemocytometer. The blood specimen is diluted using Platelet diluting fluid which preserves and fixes the platelets and stains it. The Rees – Ecker fluid is isotonic to the Platelets and doesn't cause any damage to it whereas causes the lysis of red blood cells. After diluting the specimen in appropriate dilutions, the content is charged on Hemocytometer or Neubauer's chamber and the cells are counted in the areas specific for Platelets count. Hemocytometer or Neubauer's chamber which requires experience laboratory expertise and time intensive.

1.2. LITERATURE REVIEW

[1] they have developed an automatic platelet counter for primary health care and resource-poor settings. The technology is based on a conventional microscope equipped with a digital camera linked to a personal computer, which can capture and analyze microscope images of blood samples. To evaluate the accuracy of the technology, it was compared to platelet counts performed manual by an experienced laboratory technician. Statistical analysis shows no difference between the techniques with a kappa coefficient of 0.6. This method is judged to have great potential as a tool to help primary health centers and other facilities with limited resources to deal with the burden of dengue.

[2] they proposed an image processing technique for counting the number of blood cells. The main goal of this work is the analysis and processing of a microscopic image, in order to provide an automated procedure to support the medical activity. In this project, the WBCs and RBCs are counted by using the gray thresholding algorithm computing with the manual method. This means that, the number of WBCs and RBCs are counted from the five blood images. This procedure is done because in manual counting method, the cells are counted from the five squares. After counting the number of WBCs and RBCs from these five squares, these counts are then applied to the formula to count the normalized count.

[3] they proposed which can extract platelets from the microscopic image of blood cells, and that makes platelet counting task easy. Microscopic images of stained blood slides are captured using a light microscope. Then using color-based segmentation and morphological operation, platelets can be extracted. A comparative study between the platelet counts obtained before and after segmentation along with calculation of the efficiency of the system has shown this method to be robust and effective for automation of platelet count system.

[4] Fast and cost-effective blood cell counting has major importance in the medical world. The old conventional methods of blood cell counting under the microscope render unreliable and unacceptable results and put an unendurable amount of strain on the Clinical laboratory technicians. Even so there are latest hardware solutions such as the Heska'sHema'true hematology analyzer, but they are widely unavailable and very expensive machines and so the under developed countries like Pakistan are not resourceful to provide such an expensive solution of blood cell counting in every hospital laboratory in the country. As a solution regarding this problem, this research based paper proposed a fast and cost effective software-based alternative method to count accurate blood cells.

[5] Platelet aggregation requires active platelet metabolism, platelet stimulation by agonists such as ADP, thrombin, collagen, or epinephrine; the presence of calcium or magnesium ions and specific plasma proteins such as fibrinogen or vWf; and a platelet receptor, the glycoprotein IIb/IIIa (GPIIb/IIIa) complex. Thus, platelet stimulation results in the generation of intracellular second messengers that transmit the stimulus back to the platelet surface, exposing protein binding sites on GPIIb/IIIa. Fibrinogen (or vWf) then binds to GPIIb/IIIa and crosslinks adjacent platelets to produce platelet aggregates. Platelet stimulation also results in platelet secretion and the elaboration of platelet procoagulant activity.

[6] They provide an overview of the structure of WBCs, the types and sub-types of WBC, and their features, including the shape of nuclei, size, function and colour. Next, we detail the process of the identification of WBC in images, including image acquisition and consideration of the effect of staining to visualize changes in the colour and shape of the nucleus. We then provide a survey of the recent history (since 2005) up to current state-of-the-art in automated identification of WBCs, including techniques such as image processing, signal processing, pattern recognition and deep learning techniques. We later discuss the challenges including illumination variations, changes in size and location, different maturation stages, shape, rotation, and background variations. The performance of the current techniques with respect to these challenges is evaluated. This survey will help researchers to address these challenges in future work and in the further investigation of detection, feature extraction and classification of WBCs.

[7] They presented a semi-automated tool to analyze red blood cells. We utilized circular Hough transform for circular cells detection. In a light micrograph, the red blood cells may appear in a crescent shape that known as Sicklecell. Thus, we proposed a technique to detect this shape and demonstrated its capability. Overall, our tool recorded 92% of accuracy with 70% and 72% of sensitivity and specificity respectively. Despite the promising performance, several limitations must be overcome and has been highlighted in the conclusions.

[8] they produced a survey on an image processing based system that can automatically detect and count the number of RBCs and WBCs in the blood sample image. Image Acquisition, Pre-Processing, Image Enhancement, Image Segmentation, Image Post-Processing and Counting algorithm these are six steps involved in an image processing algorithm. The objective of this research is to study the different methodologies of cells counting. Ke words: RBC, WBC, Platelets, Digital Image Processing, Morphology, Hough Transform.

[9] They proposed a method directly measures platelet function by measuring platelet adhesion to platelet-specific protein patterns using a simple, optical counting technique. The patterned surfaces are designed so that a single platelet adheres to a single protein spot and for imaging purposes, the proteins patterned on the surface are labelled with a fluorescent dye, as are the platelets that adhere to the patterned surface. Following image acquisition, the occupied (i.e. covered by a platelet) protein spots in each sample must be enumerated. The result, calculated as percent adhesion, constitutes a direct and straightforward measurement of platelet adhesion and, therefore, platelet function

[10] they prevent blood loss through processes of adhesion, activation and aggregation. Platelets play a central role in cardiovascular disease (CVD), both in the development of atherosclerosis and as the cellular mediator in the development of thrombosis. Platelets have diverse roles not limited to thrombosis/haemostasis, also being involved in many vascular

inflammatory conditions. Depending on the physiological context, platelet functions may be protective or contribute to adverse thrombotic and inflammatory outcomes. In this chapter, we will discuss platelets in context of their formation and function. Because of their multifaceted role in maintaining physiological homeostasis current and development platelet function testing platforms will be discussed

1.3. METHODOLOGY

1.3.1 IMAGE ACQUISITION OF BLOOD SMEAR

Image acquisition is the process of retrieving an image from some source, usually a hardware-based source. Digital image acquisition is done here by making use of a digital camera. A drop of blood is taken in a slide and prepares a monolayer of blood using another slice where the cells are divided sufficiently. Place a drop of immersion oil onto slide to be seen ensuring it is totally air dried subsequent to recoloring. Blend the Wright-Giemsa Stain and buffer blend to get ready with blood smears. The advanced camera is associated with a magnifying lens which catches tiny computerized picture of blood smear. Blood smear images with size of 255*255 were taken. Blood smears of normal persons and blood smears containing the dengue virus were collected and examined under a digital microscope. The entropy of the images was calculated. Entropy or average information of an image is a measure of the degree of randomness in an image. Average entropy of blood smear images taken for this work is 6.1526

1.4 DATASET

The datasets used for the Experiments and to calculate the performance evaluation are shown in Fig.1.4
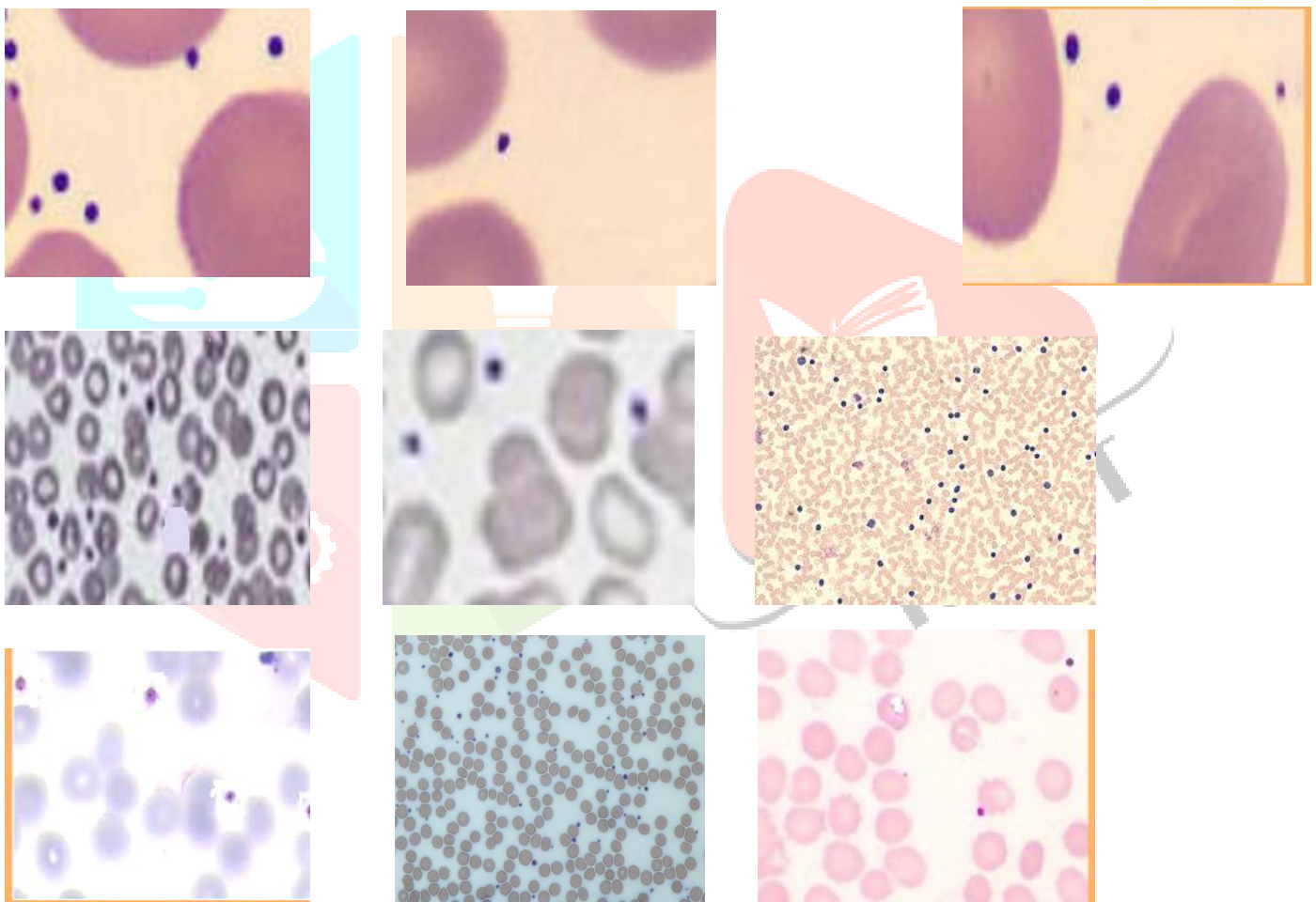
Fig 1.4 Datasets used in experiments

Blood smear of normal and dengue affected were collected for the experiment; some blood smear has wbc in it. Wbc removal is take place after the input get by mask remove method of image processing.

2. PREPROCESSING

2.1. Green channel extraction

Before segmentation the RGB color image is split into three planes such as red, green and blue. The green plane extracted is clear with feature that we need to split platelets. So the green plane of the imported image is taken for further process. The other two plane such as red and blue are not suitable for analysis with clear feature.
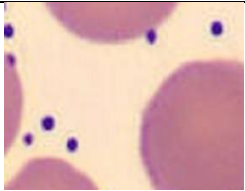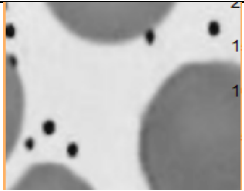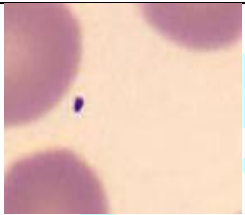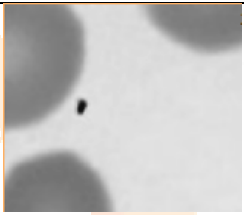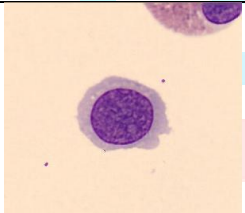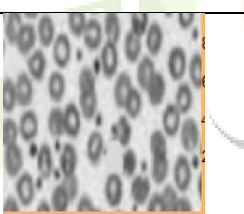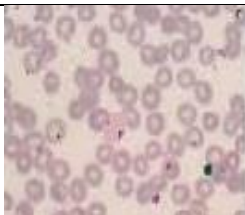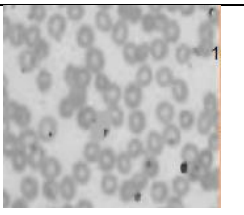
2.2. Winer filter

Filtering method has been use at this step to removing an unwanted noise, Here the wiener filter has been used for remove additive noise and restore the unclear image. The PSNR block computes the peak signal-to-noise ratio, in decibels, between two images. This ratio is often used as a quality measurement between the original and a compressed image. The higher the PSNR, the better the quality of the compress, or reconstructed image. The Mean Square Error and the Peak Signal to Noise Ratio are the two error metrics used to compare image compression quality. The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. The lower the value of MSE, the lower the error.

$$MSE = sum(sum((original\ image\text{-}filterer\ image).\text{^}2))/(M*N)$$

$$PSNR = 10*log10(256*256/MSE)$$

Table 2.2 Experimental results of Winer filter

Input Image	Output Image	PSNR	MSE
		44.02	2.0677
		48.53	0.918
		42.16	3.983
		33.566	2.88
		37.069	1.236
	Average	41.068	

Adaptive Threshold

Adaptive thresholding is used to divide desirable foreground image objects from the background based on the difference in pixel intensities of each region. This allows for thresholding of an image whose global intensity histogram doesn't contain distinctive peaks.

$$G(x,y)= 1, f(x,y) > T \quad G(x,y)= 0, f(x,y) <= T$$
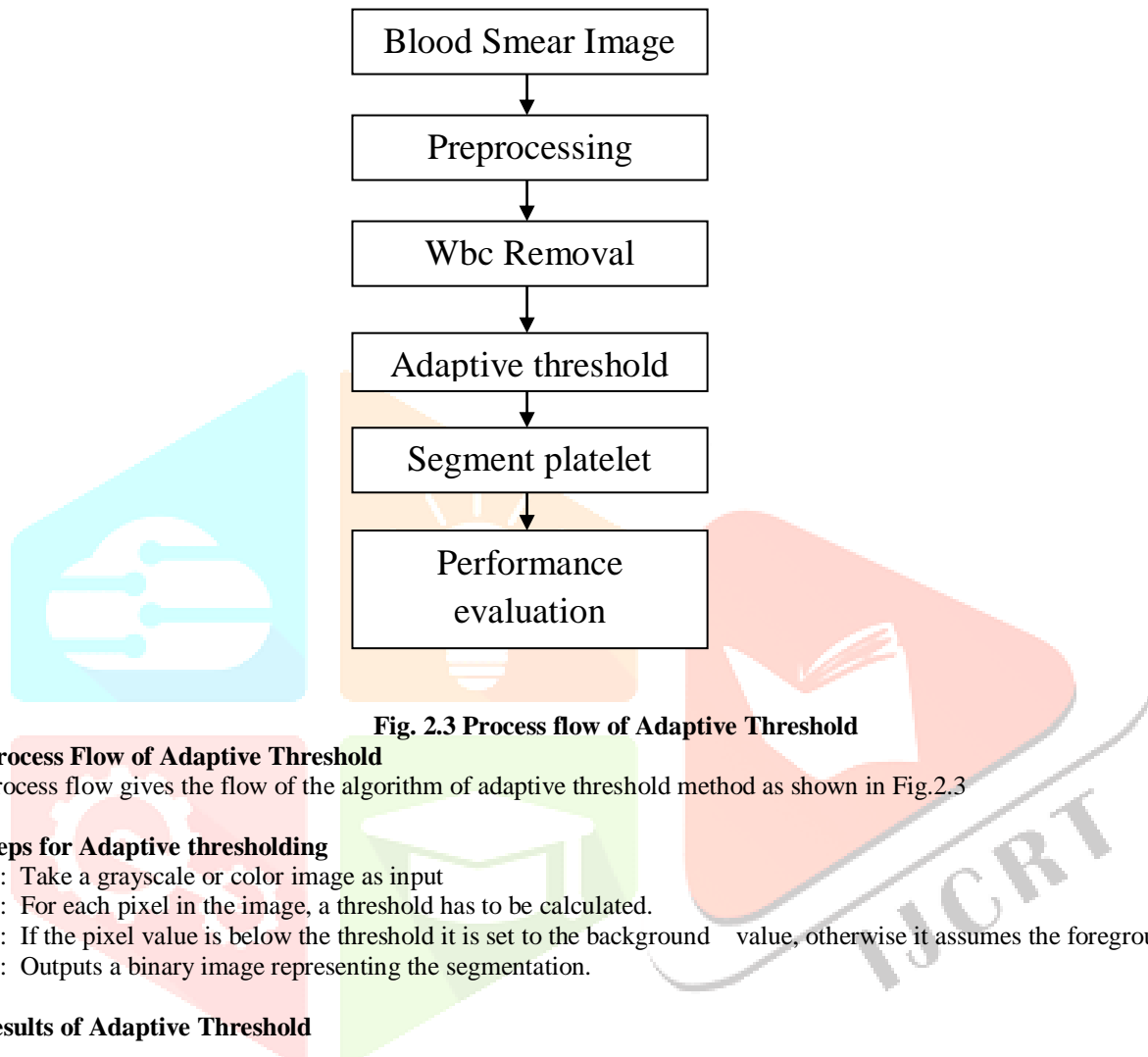
When T = T[x,y, p(x,y), f(x,y)], threshold is adaptive.

```
Blood Smear Image
       ↓
   Preprocessing
       ↓
   Wbc Removal
       ↓
 Adaptive threshold
       ↓
 Segment platelet
       ↓
   Performance
   evaluation
```

Fig. 2.3 Process flow of Adaptive Threshold

2.3. Process Flow of Adaptive Threshold
The process flow gives the flow of the algorithm of adaptive threshold method as shown in Fig.2.3

2.4 Steps for Adaptive thresholding
Step 1: Take a grayscale or color image as input
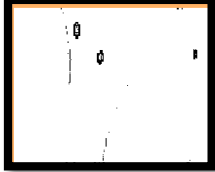Step 2: For each pixel in the image, a threshold has to be calculated.
Step 3: If the pixel value is below the threshold it is set to the background value, otherwise it assumes the foreground value.
Step 4: Outputs a binary image representing the segmentation.
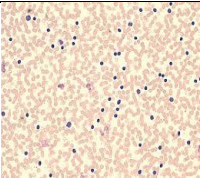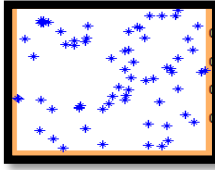
2.5 Results of Adaptive Threshold

Here various dengue affected and normal blood smear image has been segmented using the Adaptive thresholding method and the results are shown in the table 2.5

Table 2.5 Experimental Results of Adaptive Threshold

S.No	Input	Output
1		
2		
3		

2.6 Performance Analysis

Here several performance metrics are used to check the segmentation.
Segmentation results of an image and ground truth of an image are compared to evaluate the performance.
1. Accuracy
2. Precision
3. Recall

Table 2.6 Performance Evaluation Values of adaptive threshold

Image	Accuracy	Precision	Recall
1	99.4934	99.633	74.3802
2	99.6979	99.7038	89.4737
3	99.9146	99.9435	82.0755
4	98.6679	98.6677	100
5	82.2769	82.2505	96.6667
6	99.5667	99.8252	48.4848
7	99.5956	99.6681	81.9272
8	99.5811	99.7466	65.206
Average	**97.347**	**97.42**	**79.765**

FEATURE EXTRACTION AND CLASSIFICATION

3. INTRODUCTION

Feature extraction describes the relevant shape information contained in a pattern so that the task of classifying the pattern is made easy by a formal procedure. In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction.

Image classification refers to the task of extracting information classes from an image set. Depending on the interaction between the analyst and the computer during classification, there are two types of classification: supervised and unsupervised.

3.1 Feature Extraction and Classification

Platelet count and symptoms of the patient are taken as the feature.

Symptoms are fever, headache, rash, joint and muscle pain.

Table 3.1 Features Platelets and Symptoms

Count	Fever	headache	Rash	Muscle & join Pain	Class
5	1	1	1	1	1
3	1	0	0	0	2
4	1	1	1	1	1
7	0	0	1	1	2
42	1	1	1	1	2
65	0	0	0	0	2
1	1	1	1	0	1
4	1	1	1	1	1
85	0	1	1	0	2
7	1	1	1	1	1

Here symptoms are given by manual. If 0 the patient does not have that symptom otherwise 1 is entered. Two classes are separated class1 denotes affected by dengue, class 0 denotes does not affected by dengue.

3.2 MLP - MULTILAYER PERCEPTRON CLASSIFIER

Multilayer Perceptron Classifier (MLPC) is a classifier based on the feed forward. MLPC consists of multiple layers of nodes. Each layer is fully connected to the next layer in the network. Nodes in the input layer represent the input data. All other nodes maps inputs to the outputs by performing linear combination of the inputs with the node's weights w and bias b and applying activation function.

3.3 ALGORITHM

It can be written in matrix form for MLPC with K+1K+1 layers as follows:

y(x)=fK(...f2(wT2f1(wT1x+b1)+b2)...+bK)y(x)=fK(...f2(w2Tf1(w1Tx+b1)+b2)...+bK)

Nodes in intermediate layers use sigmoid (logistic) function:

f(zi)=11+e−zif(zi)=11+e−zi

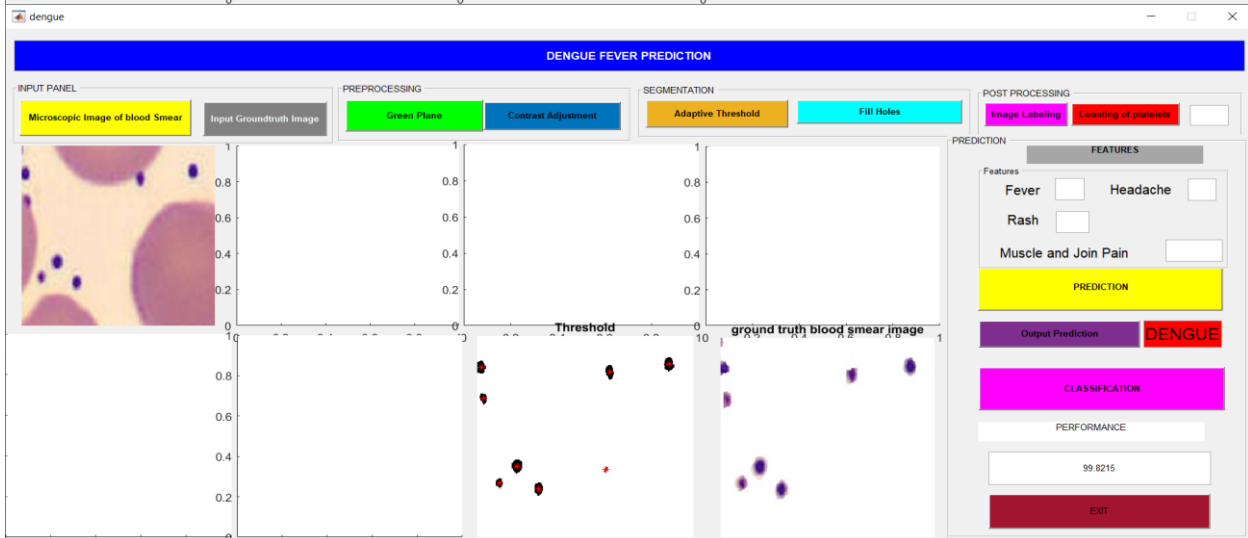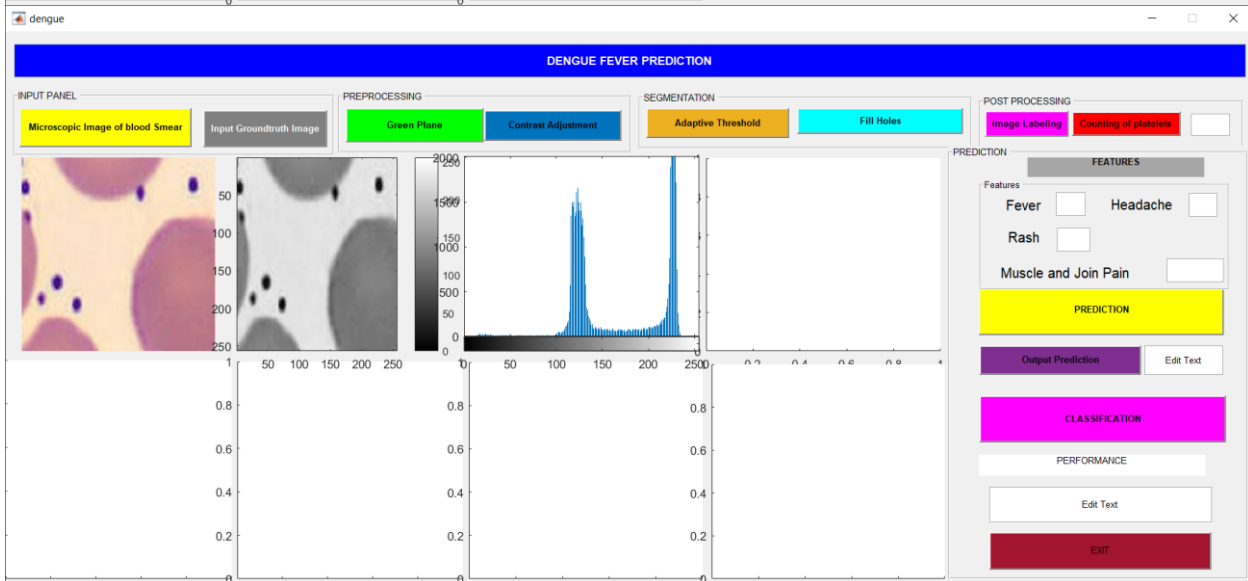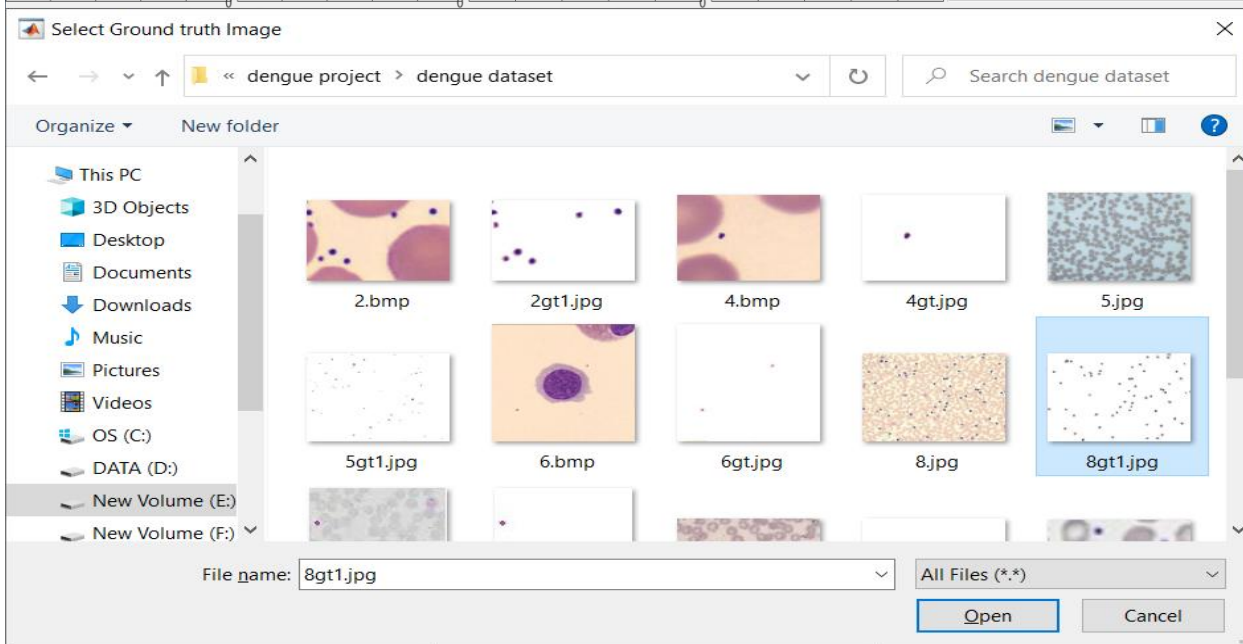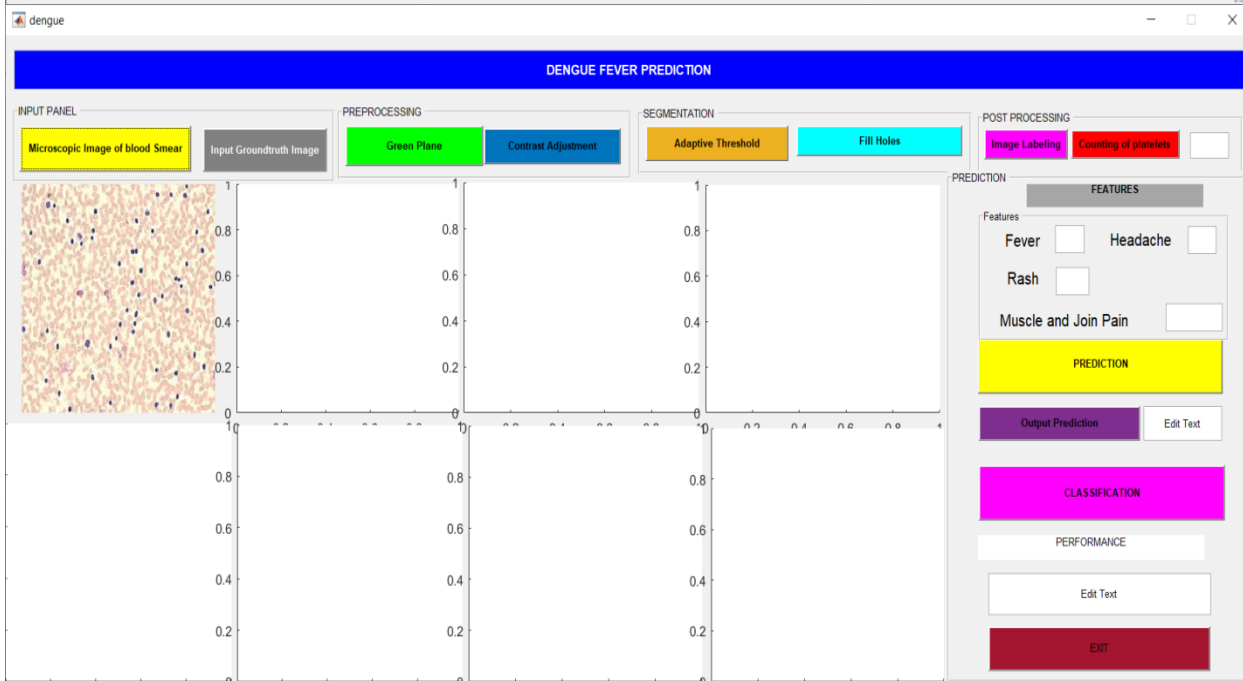Nodes in the output layer use softmax function:
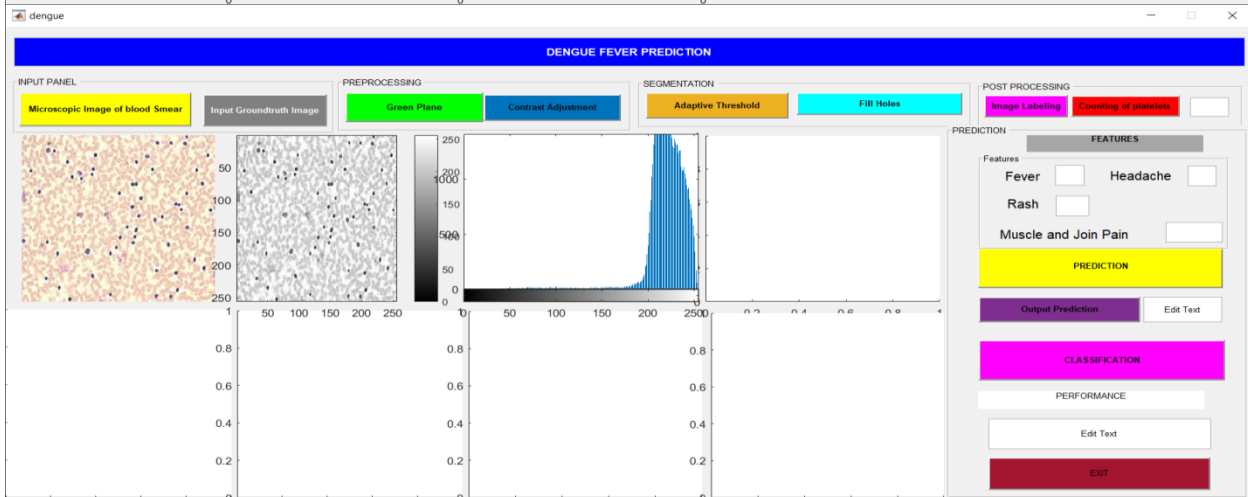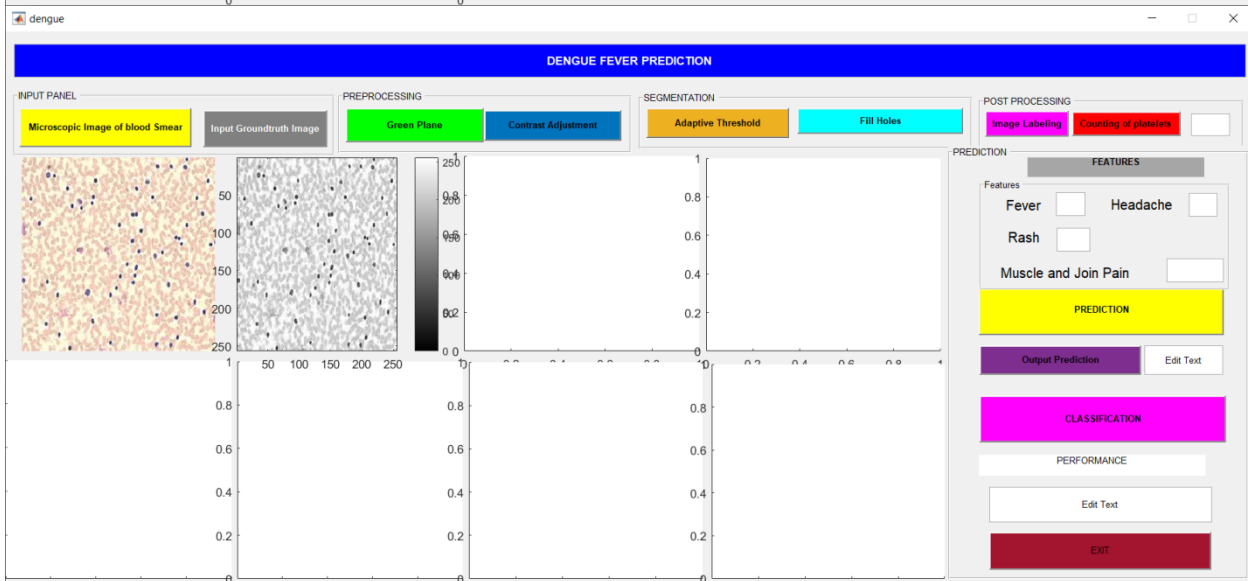
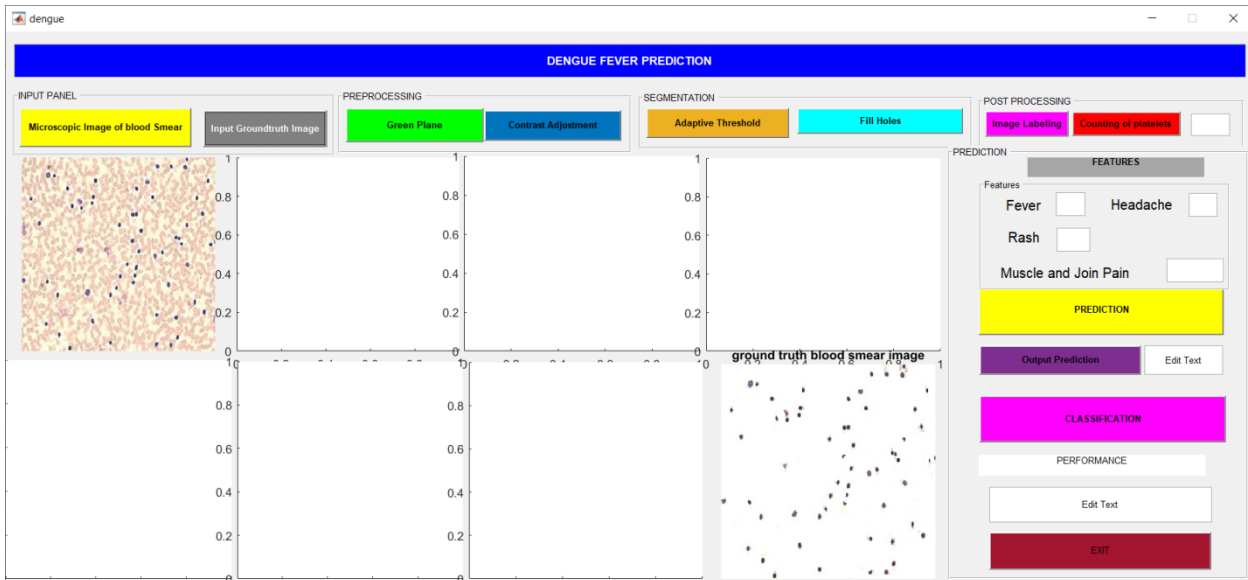f(zi)=ezi∑Nk=1ezkf(zi)=ezi∑k=1Nezk

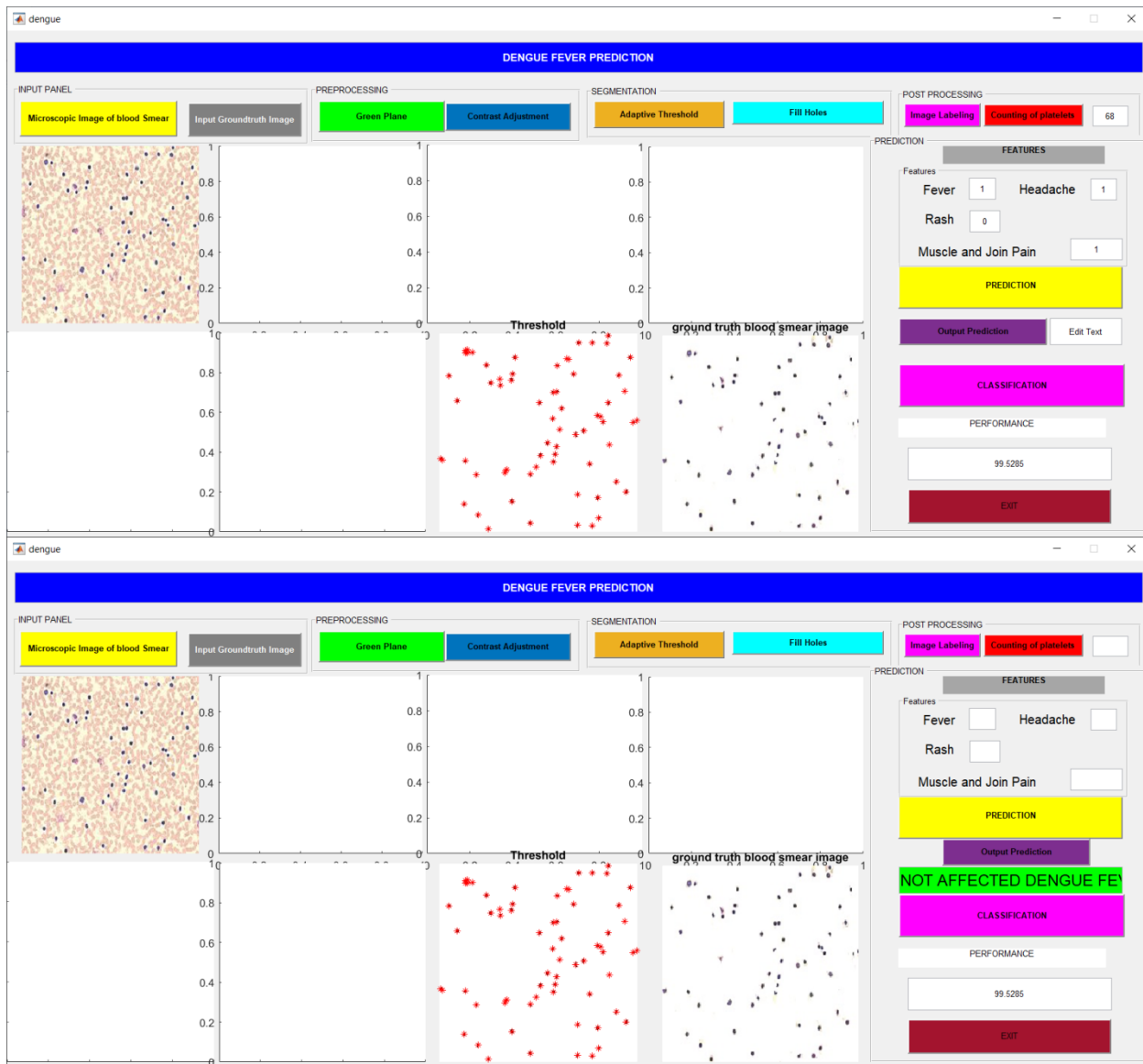The number of nodes NN in the output layer corresponds to the number of classes.

4. RESULTS
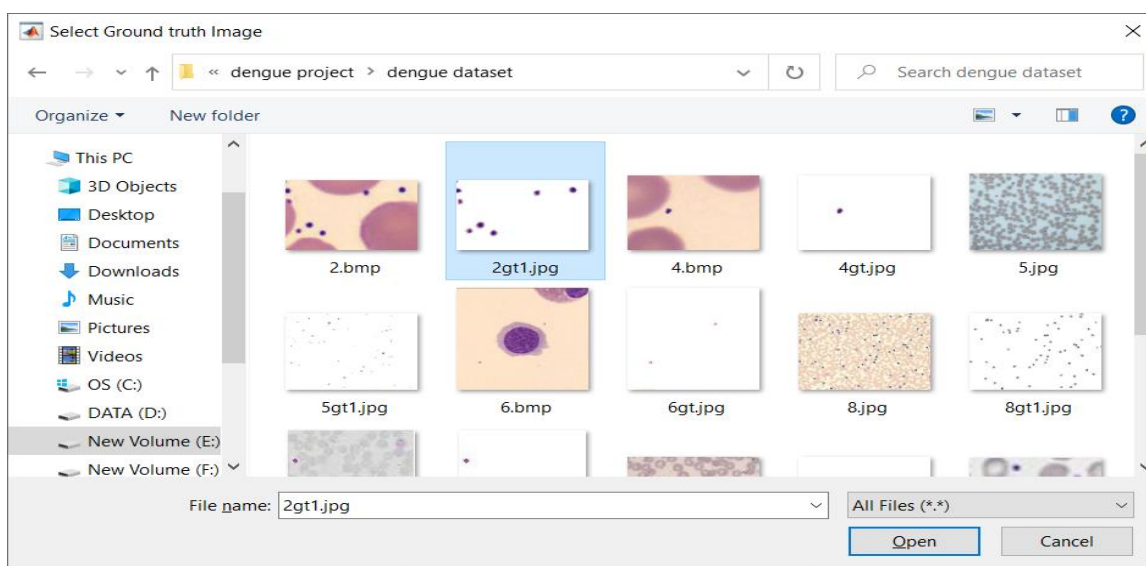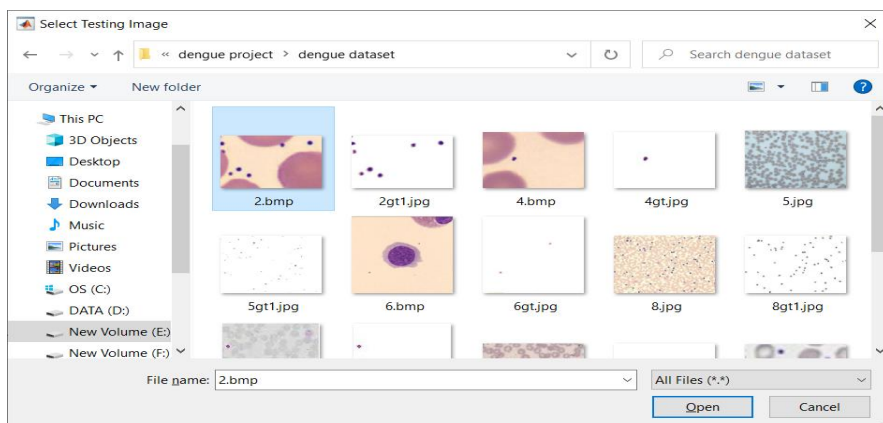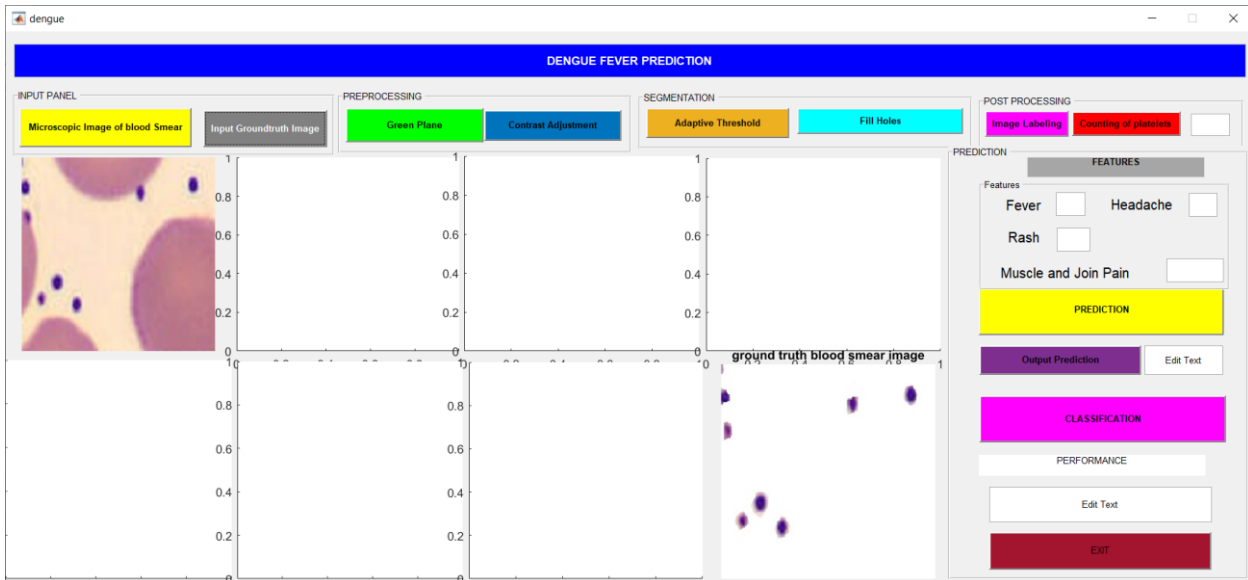
5. CONCLUSION

From the experimental results and performance evaluation global threshold segmentation method performs better than other segmentation method. Color threshold also has better performance but its recall value is low. K mean clustering method is very low performance of accuracy, precision, recall. The platelet count and symptoms are evaluated in the classification. If the platelet is below the value of ten and symptoms are marked the patient affected by dengue otherwise the patient does not affected.

6. REFERENCE

[1]Cui J, Li F, Shi ZL. Origin and evolution of pathogeni coronaviruses. Nat Rev Microbiol 2019;17(3):181–92.

[2]Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 2003;348(20):1967–76.

[3]Fouchier RA, Kuiken T, Schutten M, van Amerongen G, van Doornum GJ, van den Hoogen BG, et al. Aetiology: Koch's postulates fulfilled for SARS virus. Nature 2003;423(6937):240 .

[4]To KKW, Yip CCY, Lai CYW, Wong CKH, Ho DTY, Pang PKP, et al. Saliva as a diagnostic specimen for testing respiratory virus by a point-of-care molecular assay: a diagnostic validity study. Clin Microbiol Infect 2019;25(3):372–8 .

[5]Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020;395(10223):470–3.

[6]Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020.

[7]Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, et al. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol 2012;86(7):3995–4008.

[8]Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579(7798):265–9.

[9]Zhong NS, Zheng BJ, Li YM, Poon, Xie ZH, Chan KH, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet 2003;362(9393):1353–8.

[10]Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;579 (7798):270–3.