VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Project Report on

Detection of diseases via blood analysis using Image Processing Techniques

In partial fulfilment of the Fourth Year (Semester-VIII), Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai
Academic Year 2017-2018

Project Mentor
Ms. Kajal Jewani

Submitted by

Krishna Boddu	Roll No.11	D17C
Pratik Gurnani	Roll No.21	D17C
Kalyani Karmarkar	Roll No.27	D17C
Kiran Solapure	Roll No.60	D17C

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Certificate

This is to certify that _____ of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on ***"Detection of diseases via blood analysis using image processing techniques"*** as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Prof Ms.Kajal Jewani*** in the year 2017-2018

This thesis/dissertation/project report entitled **"Detection of diseases via blood analysis using image processing techniques"** by **Pratik Gurnani, Krishna Boddu, Kalyani Karmarkar, Kiran Solapure** is approved for the degree of _____

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5, PO6,PO7, PO8, PO9, PO10, PO11, PO12, PSO1, PSO2	

Date:

Project Guide:

Project Report Approval
For
B. E (Computer Engineering)

This thesis/dissertation/project report entitled ***"Detection of diseases via blood analysis using image processing techniques"*** by ***Pratik Gurnani, Krishna Boddu, Kalyani Karmarkar, Kiran Solapure*** is approved for the degree of

---

Internal Examiner

External Examiner

Head of the Department

Principal

Date:

Project Guide:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

----------	----------
(Signature)	(Signature)
----------	----------
(Name of Student and Roll No.)	(Name of Student and Roll No.)
----------	----------
(Signature)	(Signature)
----------	----------
(Name of Student and Roll No.)	(Name of Student and Roll No.)

Date:

Acknowledgments

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Ms. Kajal Jewani** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair** , for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO 2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solution for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen ones competency.
CO 11	Able to develop professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Abstract

Blood related diseases like Malaria and Acute Lymphoblastic Leukemia are responsible for the deaths of millions of people each year. Early diagnosis of the disease is necessary for their correct identification and treatment. Malaria and Acute Leukemia are diagnosed by drawing blood sample from the patient's body and observing the thin blood smear under the microscope to check for irregularities. This requires skill and expertise and is prone to human error. The proposed method constitutes an android application which acts as a portable and inexpensive means of diagnosis via image processing and analysis.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Nearly half of the world's population is at risk of malaria. In 2015, there were roughly 212 million malaria cases and an estimated 429 000 malaria deaths.[1] Approximately every 3 minutes one person in the United States (US) is diagnosed with a blood cancer. An estimated combined total of 172,910 people in the US are expected to be diagnosed with leukemia in 2017.[2] Delay in the diagnosis or failure to detect diseases at early stages leads to further complications, and easily preventable deaths. Recovery is only possible if one can get sure & accurate analysis at an early stage.

1.2 Problem Definition

Manual detection of diseases via microscopic image examination is prone to human error, and requires implicit trust in pathologist expertise. Another major roadblock is the unavailability of modern medical facilities in rural areas. We aim to develop a system that is simplistic, accessible, requires fewer resources while simultaneously providing faster, but as accurate diagnoses.

1.3 Relevance Of The Project

In contrast to traditional methods, the proposed method ensures efficiency by diagnosing the disease digitally. Since there is no reliance on human pathologists, the rate of diagnosis can theoretically amplify to rate of smartphone availability. Combined with lowering prices of medicines, we could potentially eradicate some of the most common diseases of our era. As smart-phones become ubiquitous even in rural areas, ordinary people can use this application anywhere and any time at zero cost.

1.4 Methodology Used

Our goal is to develop a system for detection of specific irregularities among the constituent blood cells in a microscopic image for diagnosis of diseases like Malaria and Leukemia, all of which follow a similar sequence of events for diagnosis. We have used iterative approach for developing our system.

System involves following steps for detection of :

1.4.1 Malaria

Image Acquisition

The image can be obtained from either an existing dataset for testing, or can be from a drawn blood sample. To obtain a microscopic image of the slide, we could use a microscope with a camera attachment, or alternatively, use a device like Foldscope[3], an ultra-low-cost optical microscope with a smartphone attachment. This involves conversion of microscopic images into digital format. Input images of stained blood smears are selected from the dataset library. Dataset consists of infected and healthy sample images.

Take Image Through Phone/Gallery

In the proposed system, user can take microscopic image from camera as well as from gallery. Further, this image will be uploaded to the server where actual processing will take place.

Image Pre-Processing

Microscopic input image is converted from RGB to Grayscale to reduce pre-processing time. RGB to Gray conversion is done by averaging all three components i.e. R, G and B which results in Grayscale. Pre-processing removes unwanted effects from the image and adjust the image for further processing. This step also involves image augmentation. After flipping and rotating the same image below results are obtained:

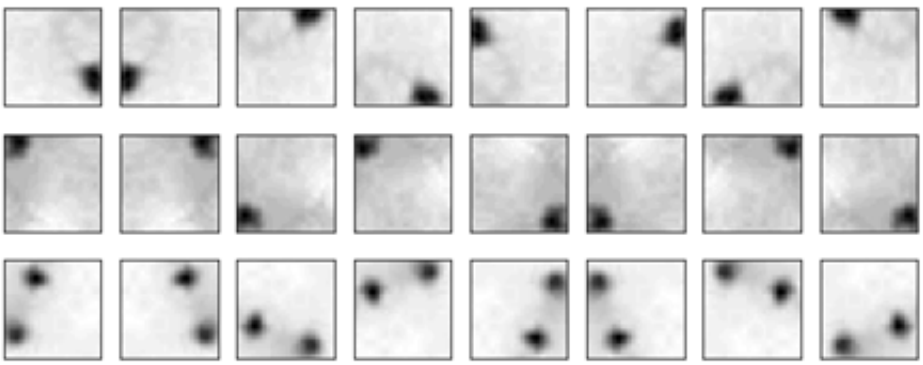


Figure 1.1: Image Augmentation

Classification Method

KNN Classification: K-Nearest Neighbor classifier is a machine learning algorithm which classifies datasets based on their similarity with neighbours.It is a non-parametric classifier, i.e.it does not make any underlying assumptions about the distribution of data. It is also considered to be an instance based learning/non-generalizing algorithm. Training process for this algorithm only consists of storing feature vectors and labels of the training images. In the classification process, the unlabeled query point is simply assigned to the label of its k-nearest neighbors. Typically, the class of the object is decided on the basis of its k-nearest by majority votes. If k=1, object is classified as the class of the object nearest to it.In case of only two classes, k must be an odd integer. However, there can still be ties when k is an odd integer while performing multi class classification. Hence, Euclidean distance is used to find the distance which is most common function for KNN.

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^{m} (x_i - y_i)} \tag{1.1}$$

ALGORITHM

- Let *m* be the number of training data samples. Let *P* be an unknown point.
- Store the training samples in an array of data points *arr[]* .This means each element of this array represents a tuple *(x, y)*.
- For *i=0 to m*, calculate Euclidean distance *d(arr[i], p)*.
- Make set *S* of *K* smallest distances obtained. Each of these distances corresponds to an already classified data point.
- Return the majority label among *S*.

Result

From the server, the obtained result will be transferred and displayed below the image in the application. Final diagnosis result can be seen in the application or on the server side also. It directly displays malaria positive or negative.

1.4.2 Leukemia

Image Acquisition

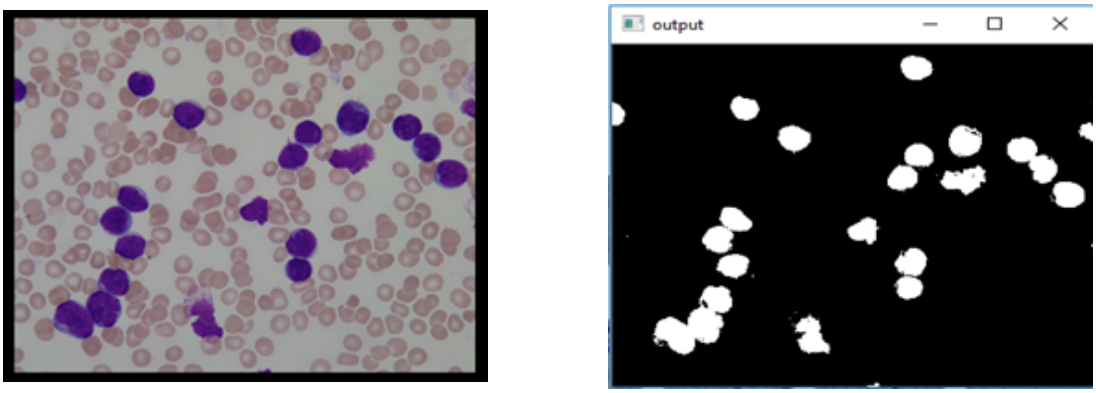
The image can be obtained from either an existing dataset for testing, or can be from a drawn blood sample. To obtain a microscopic image of the slide, we could use a microscope with a camera attachment, or alternatively, use a device like Foldscope[3], an ultra-low-cost optical microscope with a smartphone attachment. This involves conversion of microscopic images into digital format. Input images of stained blood smears are selected from the dataset library. Dataset consists of infected and healthy sample images.

Take Image Through Phone/Gallery

In the proposed system, user can take microscopic image from camera as well as from gallery. Further, this image will be uploaded to the server where actual processing will take place.

Image Processing

In this step, the image is read by the system, and this RGB image is converted to HSV. Now we segment the image based on the color of the dye by finding the maximum and minimum value of the color of the dye. We use the inrange function of OpenCV that does the work of segmenting the image with white pixels as infected cells.



(a) Input Image

(b) After pre-processing

Figure 1.2: Leukemia Image Pre-Processing

Pixel Counting

After segmenting the image, we find the white pixel counts. In case of infected sample the number of WBCs is huge in number whereas in case of healthy sample the number of WBCs is very less. So the pixel counts for infected sample is huge whereas for uninfected sample it is less. Based on the dataset available, we can make a rough estimation of the pixel count which act as a threshold by considering True Positive, False Positive, True Negative, False Negative. In our case, we have kept a threshold value of pixel count as 5386 i.e. if the pixel count exceeds 5386 then the patient surely has leukemia, and if the patient has count less than 3011 he doesnt have leukemia.

ALGORITHM

- Find min and max values of the colour of the dye used for the infected cells.
- Segment the image on the basis of the min, max value of the colour of the dye
- Count(c) the number of white pixels(infected cells) in the image.
- If(c $<$3011) then print Leukemia Negative
- If(c $>$5386) then print Leukemia Positive
- Else print May Or May Not Have Leukemia

Result

From the server, the obtained result will be transferred and displayed below the image in the application. Final diagnosis result can be seen in the application or on the server side also. It directly displays Leukemia positive or negative.

Chapter 2

Literature Survey

2.1 Research Papers

TITLE: Detection of malaria parasite species and life cycle stages using microscopic images of thin blood smear, some more text

AUTHORS: Akshay Nanoti, Sparsh Jain, Chetan Gupta, Garima Vyas

ABSTRACT:

Malaria is responsible for nearly 438,000 deaths worldwide in a year. A total of 214 million cases of malaria are encountered annually. The conventional method for testing malaria is through microscopy. A blood sample of the patient is spread over a glass slide, stained with Giemsa stain and examined under a microscope. It takes a few hours and a highly trained professional to visually examine the slide and give the results. It is even more difficult to detect the different types of malaria parasite and their stages by the conventional methods. The proposed method involves acquisition of the thin blood smear microscopic image at 100x magnification, pre-processing by partial contrast stretching, separation of infected cell from the image by applying k-means clustering on the a*b component of L*a*b colour space, feature extraction (shape and textural) of the infected cell, feature reduction using one way ANOVA and finally training the K-nearest neighbour classifier to test the images. Instead of extracting features for the entire group of erythrocytes present in the image, the algorithm only processes the infected cells increasing the speed, effectiveness and efficiency of testing. The KNN classifier is trained with 300 images to detect three lifecycle stages (trophozite, schizont and gametocyte) for each of the four species of malarial parasites (P.falciparum, P.vivax, P.malariae, and P.ovale) with an accuracy of 90.17% and sensitivity of 90.23%.

INFERENCE:

Detection of malaria parasite specie and its life cycle stage requires a highly trained pathologist, as there exist a large number of variable features making it difficult to distinguish these parasites. The segmentation of malarial parasite by k-means clustering applied to the a*b component of the L*a*b colour space helps in improving the efficiency of the algorithm as features are extracted for only the parasites rather than the entire group of erythrocytes present. The method proposed helps in detection and classification of malarial parasites and there life cycle stages with an accuracy of 90.17% and sensitivity of 90.23%.

TITLE: Detection of presence of Parasites in Human RBC. In Case of Diagnosing Malaria Using Image Processing

AUTHORS: Pranati Rakshit, Kriti Bhowmik

ABSTRACT:

Malaria is the commonest protozoal infestation in human being residing in nearly 3 billion victims across 107 countries and 1-3 million deaths per year round the globe. The disease is generally diagnosed by examining properly stained peripheral blood smear as the malarial parasite particularly invades red blood corpuscles (RBC) of the circulatory system. For this reason, proper analysis of RBC is the most confirmatory diagnosis of malaria. Here in this paper, correct identification of presence of malarial parasite within RBC has been detected and severity of the disease is measured by analysing the stage (i.e. Ring trophozoite, Merozoite, Schizont etc.) of Plasmodium sp., the malarial parasite using different image processing tools and techniques. After several pre-processing activities, area of the infested corpuscle is calculated and Sobel Edge detection method is used to find the boundary of the corpuscles. Then Harris corner points are used to formulate a metric that can determine the severity of the disease. The purpose of this paper is to highlight this medico-technical aspect only.

INFERENCE:

This present piece of work is innovative of its kind where some infected blood samples are analysed to determine stage of the disease malaria. Apart from detecting the presence of diseased RBCs in the blood smear, it has been seen that the value of the metric we proposed here is lower for earlier stages of malaria and it increases with the maturity of malarial parasite within RBC. Thus apart from detecting malaria, severity of the disease in human body can also be diagnosed in an efficient way through this technique.

TITLE: Automatic Detection Of Malaria Parasite From Blood Images

AUTHORS: Ms. Deepali Ghate,. Mrs. Chaya Jadhav, Dr. N Usha Rani

ABSTRACT:

Malaria is a serious disease for which the immediate diagnosis is required in order to control it. Microscopes are used to detect the disease and pathologists use the manual method due to which there is a lot of possibility of false detection being made about the disease. If the wrong detection is done then the disease can turn into more severe state. So the study about the computerized diagnosis is done in this paper, which will help in immediate detection of the disease to some extent, So that the proper treatment can be provided to the malaria patient. Also the image processing algorithm is used which will reliably detect the presence of malaria parasite from Plasmodium falciparum species in thin smears of Giemsa stained peripheral blood sample. Some image processing algorithms to automate the diagnosis of malaria on thin blood smears are developed, but the percentage of parasitaemia is often not as precise as manual count. One reason resulting in this error is ignoring the cells at the borders of images. This paper removes the human error while detecting the presence of malaria parasites in the blood sample by using image processing and automation. This is achieved by using Image Segmentation techniques to

detect malaria parasites in images acquired from Giemsa stained peripheral blood samples. This is comparative study of two methods for detecting malaria parasites, first method is based on segmentation and second uses feature extraction using minimum distance classifiers. We built the malaria detection system in a robust manner so that it is unaffected by the exceptional conditions and achieved high percentages of sensitivity, specificity, positive prediction and negative prediction values.

INFERENCE:

The detection of Malaria parasites is done by pathologists manually using microscopes. So, the chances of false detection due to human error are high, which in turn can result into fatal condition. This paper curbs the human error while detecting the presence of malaria parasites in the blood sample by using image segmentation and feature extraction using minimum distance classifier. It shows the comparative study between two methods as mentioned above. In image segmentation we are getting the accurate and required results in the short period of time whereas in case of feature extraction more time is required i.e more CPU utilization is there.. The system in a robust manner so that it is unaffected by the exceptional conditions and achieved high percentages of sensitivity, specificity, positive prediction and negative prediction values. And the extraction of red blood cells achieves a reliable performance and the actual classification of infected cells.

TITLE: Detection and Counting of Blood Cells in Blood Smear Image

AUTHORS: K.Pradeep , C.Ganthimathi , K.Harini and N.Diddha

ABSTRACT:

This paper deals with an image processing technique used for detecting the blood cells in less time. The proposed technique also helps in counting and segregating the blood cells in blood smear image of different categories based on the form factor using various Morphological operations. Nowadays in Hospitals and clinical Laboratories the waiting time for getting their blood results and reports are more commonly 24 hours to 8 days in case of high severity diseases where the mortality rates are high. Doctors and technicians in healthcare sectors recommended that the patients waiting time should be as less as possible and the treatment should be started immediately for the high risk diseases like Hepatitis B. The major other factor affects patient in healthcare field is the more expensive pathological tests which sometimes leading to loss of patients life. The proposed technique gives improved accuracy in counting the number of blood cells in blood smear image in compare to manual counting in laboratories.

INFERENCE:

As a conclusion, this proposed technique successfully utilizes morphological approached for segmentation, extraction and estimation in order to solve problem in image processing of the blood cells. The results of the image can be used as good input in detecting the number of blood cells in blood smear image. With framed algorithm, the blood cells can be detected and segmented as well as estimated the number of the blood cells for counting. Through system created using MATLAB, it also enable the study of the morphological features of the blood

cells image, thus, can determine whether the person is normal or otherwise by referring amount of blood cells in human blood. This technique does not involve too much looping process when develops the MATLAB source code program. One of the factors that need to be considered in healthcare field to improve this study is to reduce the time taken by the user determine the blood cells parameters in laboratories and hospitals

TITLE: Deep Convolutional Neural Networks for Microscopy-Based Point of Care Diagnostics

AUTHORS: John A. Quinn , Rose Nakasi, Pius K. B., Patrick Byanyima, William Lubega, Alfred Andama

ABSTRACT:

Point of care diagnostics using microscopy and computer vision methods have been applied to a number of practical problems, and are particularly relevant to low-income, high disease burden areas. However, this is subject to the limitations in sensitivity and specificity of the computer vision methods used. In general, deep learning has recently revolutionised the field of computer vision, in some cases surpassing human performance for other object recognition tasks. In this paper, we evaluate the performance of deep convolutional neural networks on three different microscopy tasks: diagnosis of malaria in thick blood smears, tuberculosis in sputum samples, and intestinal parasite eggs in stool samples. In all cases accuracy is very high and substantially better than an alternative approach more representative of traditional medical imaging techniques.

INFERENCE:

In this paper we have shown that the performance improvement by deep learning and CNNs compared to alternatives in other application domains can successfully be translated to point of care microscopy-based laboratory diagnosis. In contrast to systems with hand engineered features for each problem, in this case the method learns good representations of data directly from the pixel data. The fact that in our experiments the same network architecture successfully identifies objects in three different types of sample further indicates its flexibility; better results still are likely given task-specific tuning of model parameters with cross validation for each case. This improvement in performance can advance microscopybased POC diagnostics which is particularly relevant in the developing world where both microscopes and smartphones are more readily available than skilled laboratory staff. Even where laboratory staff are present in this context, this type of system can be utilised as a decision support tool, identifying possible pathogens in an image, with the technician making the final decision. This mode of use can help laboratory staff to achieve consistency in diagnosis, and by focusing concentration on parts of the images likely to contain pathogens, may also help to relieve operator fatigue and improve throughput rates.

TITLE: Detection of Leukemia in microscopic images using image processing

AUTHORS: Chaitali Raje,Jyoti Rangole

ABSTRACT:

Leukemia occurs when lot of abnormal white blood cells produced by the bone marrow. Hematologist makes use of microscopic study of human blood, which leads to need of methods, including microscopic color imaging, segmentation, classification and clustering that can allow identification of patients suffering from Leukemia. The microscopic images will be inspected visually by hematologists and the process is time consuming and tiring. The automatic image processing system is urgently needed and can overcome related constraints in visual inspection. The proposed system will be on microscopic images to detect Leukemia. The early and fast identification of Leukemia greatly aids in providing the appropriate treatment. Initial segmentation is done using Statistical parameters such as mean, standard deviation which segregates white blood cells from other blood components i.e. erythrocytes and platelets. Geometrical features such as area, perimeter of the white blood cell nucleusis investigated for diagnostic prediction of Leukemia. The proposed method is successfully applied to a large number of images, showing promising results for varying image quality. Different image processing algorithms such as Image Enhancement, Thresholding, Mathematical morphology and Labelling are implemented using LabVIEW and MATLAB.

INFERENCE:

The main aim of this paper is nucleus segmentation followed by feature extraction to detect Leukemia. Shape features of nuclei such as area, perimeter, etc. are considered for better accuracy of detection. The results show that the proposed statistical parameter such as mean and standard deviation based image segmentation and Otsu's thresholding based produced good segmentation performance. In addition, the fully segmented nucleus can be better achieved by using LabVIEW based algorithm because it is less sensitive to input image variations.

Chapter 3

Requirements

3.1 Functional Requirements

1. There should be true colour input image.
2. Application should be able to handle various image file types.
3. The blood cells should be clearly observable.
4. The slide should not have overlapping layers of cells.
5. Datasets of the blood borne diseases has to be known.
6. We must know the location of hospitals near to the user.

3.2 Non-Functional Requirements

1. Ability to take input images from multiple sources like:
    - Dataset
    - Microscope with Camera
    - Phone with Foldscope attachment
2. Processing time should be low.
3. Quality of the image should not be compromised.
4. The proposed technique should ensure maximum accuracy.
5. It should be accessible to everyone regardless of any expertise.

3.3 Constraints

1. Image quality should not affect the system
2. The system should maintain reasonable performance standards and be secure towards individuals data

3. The system should be implemented in given time frame.
4. Transfer of image from user to server should be within seconds.
5. Nearby location should be shown dynamically.

3.4 Hardware And Software Specifications

3.4.1 Hardware Specifications

1. Smartphone
2. Computer for processing (as Server)
3. Microscope with camera attachment (Optional)
4. Paper microscope that provides the magnification from 140X to 2000X. (Optional)

3.4.2 Software Specifications

1. Supported in Windows 7, 8, 8.1, 10 (32 or 64 Bit OS)
2. Supported in all the Android Versions above Android 2.3(Gingerbread)

Chapter 4

Proposed Design

4.1 Block Diagrams

4.1.1 Malaria

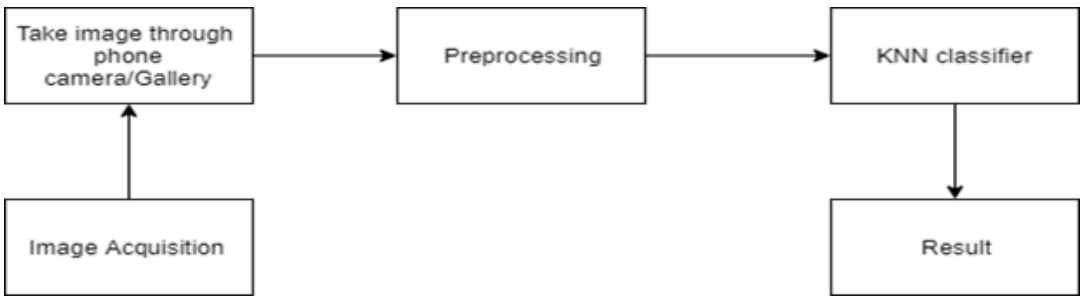


Figure 4.1: Block Diagram For Malaria

4.1.2 Leukemia

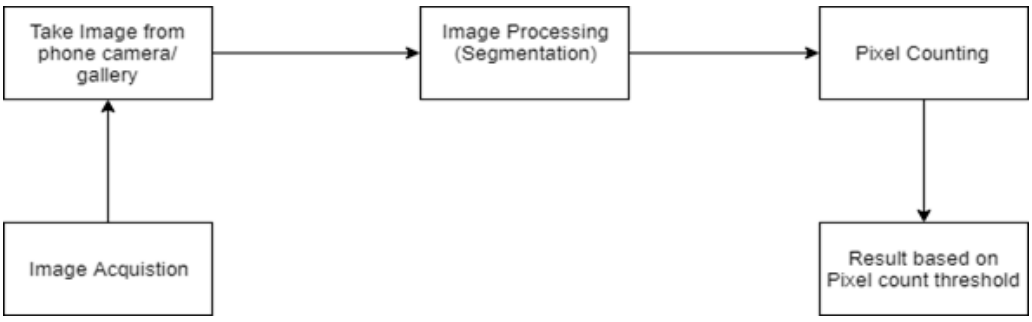


Figure 4.2: Block Diagram For Leukemia

4.2 Detailed Design

4.2.1 Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the "flow of data through an information system, modelling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design). A DFD shows what kind of information will be input to and output from the system, how the data will advance through the system, and where the data will be stored. It does not show information about process timing or whether processes will operate in sequence or in parallel, unlike a traditional structured flowchart which focuses on control flow, or a UML activity workflow diagram, which presents both control and data, flows as a unified model.

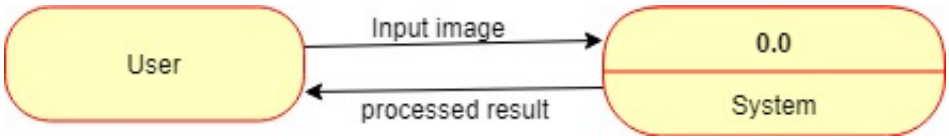


Figure 4.3: DFD Level 0

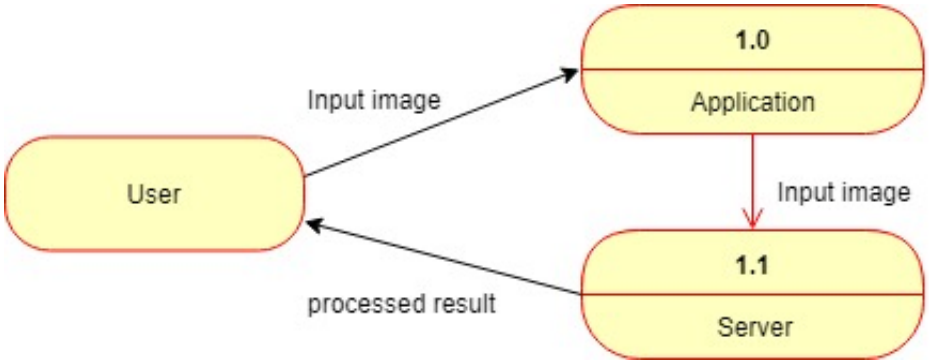


Figure 4.4: DFD Level 1

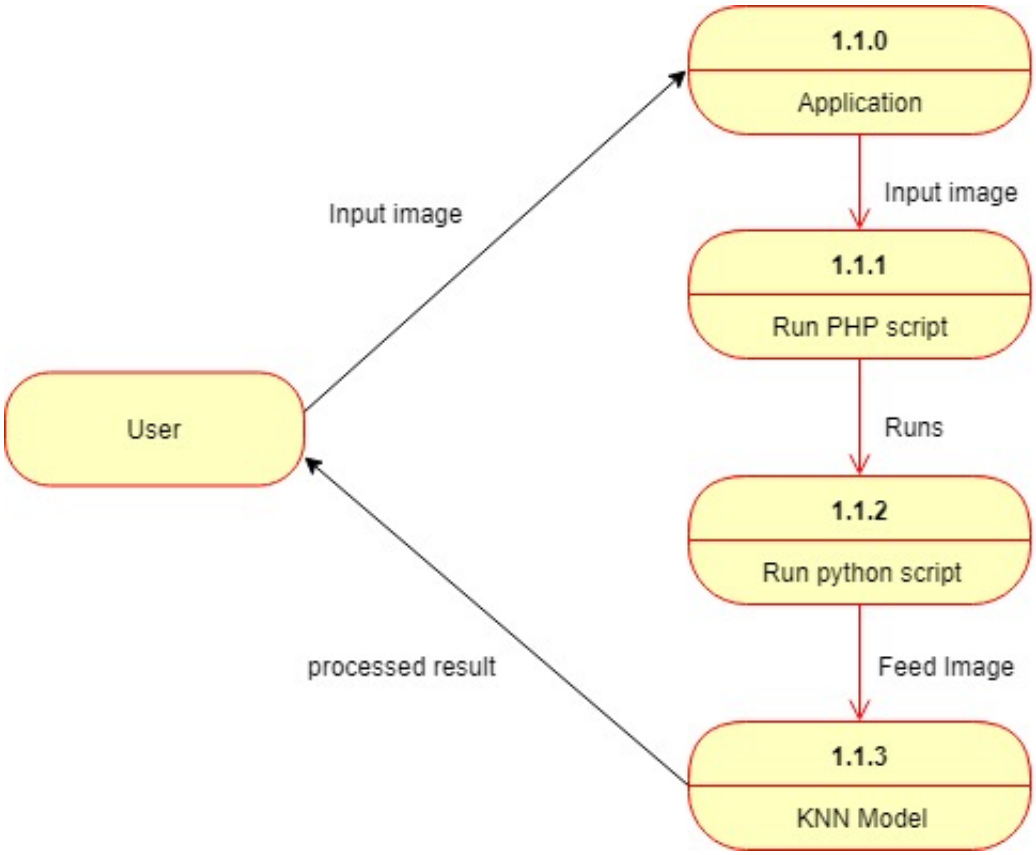


Figure 4.5: DFD Level 2

4.2.2 Use Case Diagram

Its purpose is to represent a graphical overview of the functionality provided by the system in terms of actors, their goals(represented as a use cases), and any dependencies between those use cases. As shown in the use case diagram, two actors user and developer interact with each other through use cases after acquiring image from blood sample, image acquisition is done after which developer does pre-processing, image analysis, diagnosis and final result is send to the user thereby providing nearby locations to the user in the app. Final result and depending on the result nearby locations will be shown to the user in the app.

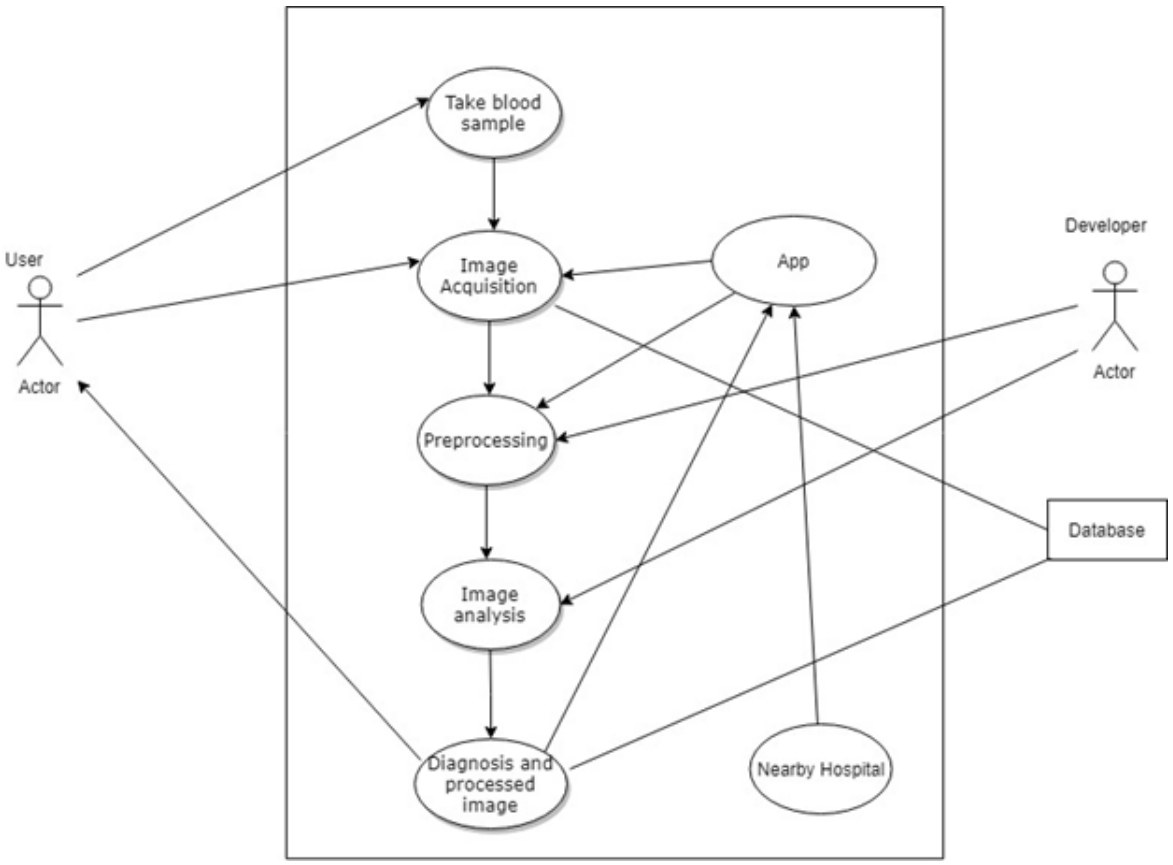


Figure 4.6: Use Case Diagram

4.2.3 Modular Diagram

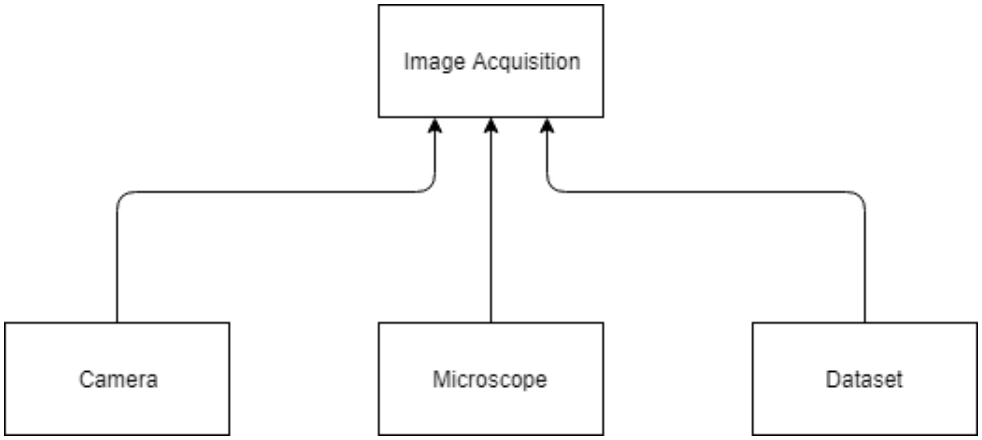


Figure 4.7: Image Acquisition Module Diagram

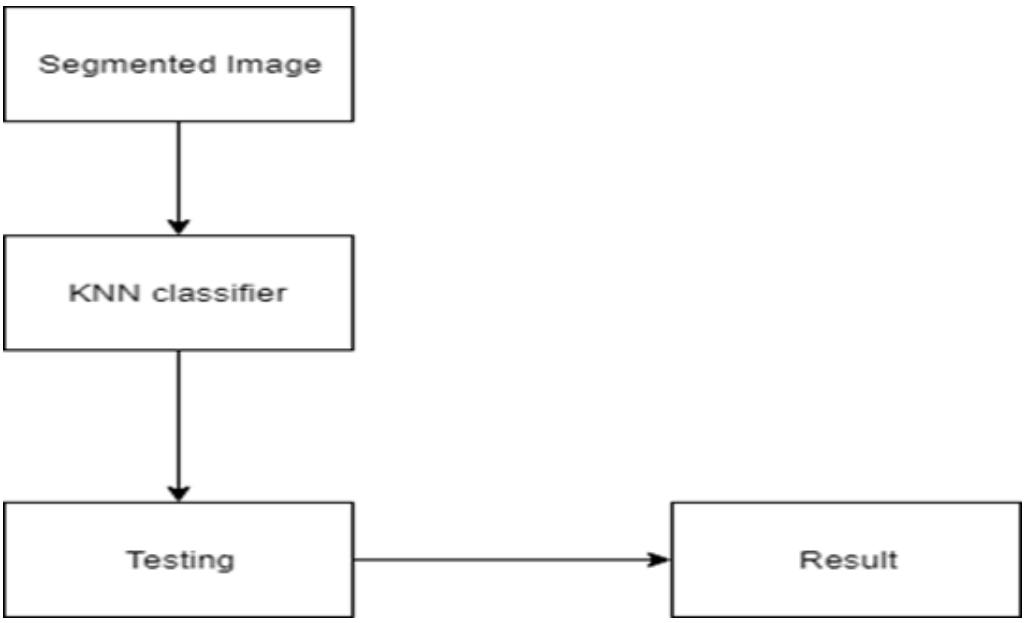


Figure 4.8: Malaria Module Diagram

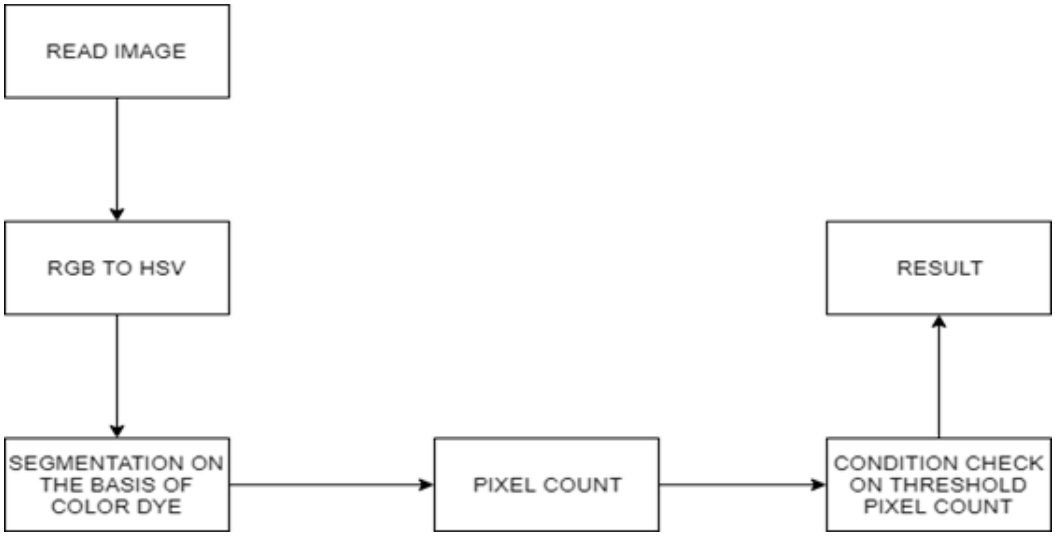


Figure 4.9: Leukemia Module Diagram

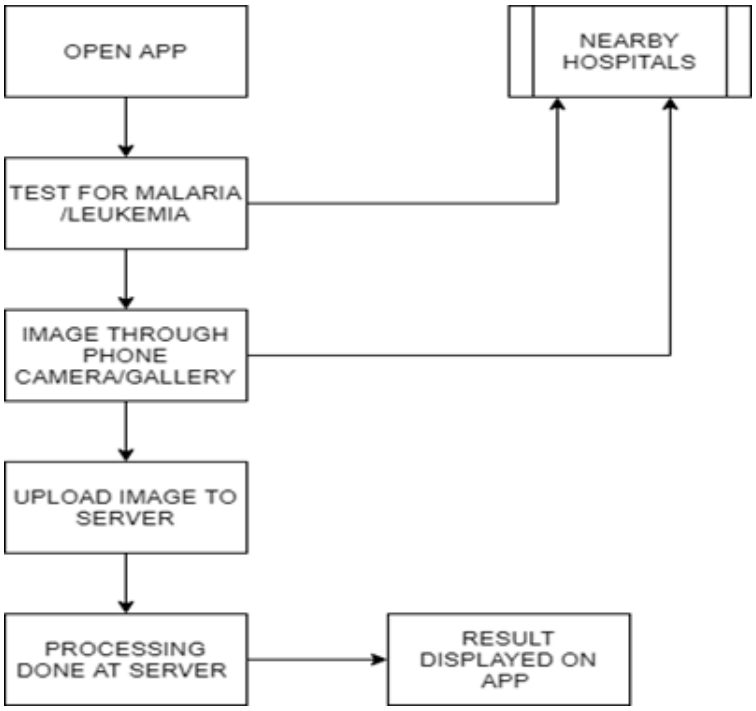


Figure 4.10: Android App Module Diagram

Chapter 5

Implementation

5.1 Code for disease diagnosis

PHP script to run python code

```
<?php

// Path to move uploaded files
$target_path = dirname(__FILE__).'/uploads/';

if (isset($_FILES['image']['name'])) {

   $target_path = $target_path . basename($_FILES['image']['name']);
      //echo $target_path
   try {
      // Throws exception incase file is not being moved
      if (!move_uploaded_file($_FILES['image']['tmp_name'],
         $target_path)) {
         // make error flag true
         echo json_encode(array('status'=>'fail', 'message'=>'could
            not move file'));
      }

      // File successfully uploaded
      //echo json_encode(array('status'=>'success', 'message'=>'File
         Uploaded'));

            ini_set('max_execution_time', 300);
            $python = "C:\\Users\\krishna\\AppData\\Local\\Programs\\
               Python\\Python35\\python.exe";
            $file = "C:\\xampp\\htdocs\\ImageUpload\\python\\test.py
               ";
            $output=exec($python . " " . $file);
            echo json_encode(array('status'=>'success', 'message'=>
               $output));

   } catch (Exception $e) {
      // Exception occurred. Make error flag true
```

```
        echo json_encode(array('status'=>'fail', 'message'=>$e->
            getMessage()));
    }
} else {
    // File parameter is missing
    echo json_encode(array('status'=>'fail', 'message'=>'Not received
        any file'));
}

?>
```

Python script to check for malaria

```
import numpy as np
import re
import os
from sklearn.externals import joblib
import matplotlib.image as mpimg

def rgb2gray(rgb):
   return np.dot(rgb[...,:3], [0.299, 0.587, 0.114])

directory = 'C://xampp//htdocs//ImageUpload//uploads'
for file in os.listdir(directory):
   filename = os.fsdecode(file)

   if filename.endswith(".jpg"):
      clf = joblib.load('C:\\xampp\\htdocs\\ImageUpload\\python\\
         knn10.pkl')
      img = directory+'//'+str(filename)
      #print(img)
      test_img = mpimg.imread(img)
      gray = rgb2gray(test_img)
      y_pred = clf.predict_proba(gray.flatten().reshape(1, -1))
      #print(y_pred)
      #print(np.argmax(y_pred))
      #print(y_test[1])
      if(np.argmax(y_pred)) == 0:
         print('Malaria Negative')
      else:
         print('Malaria Positive')
      # index +=1
      os.remove(img)
   else:
      continue
```

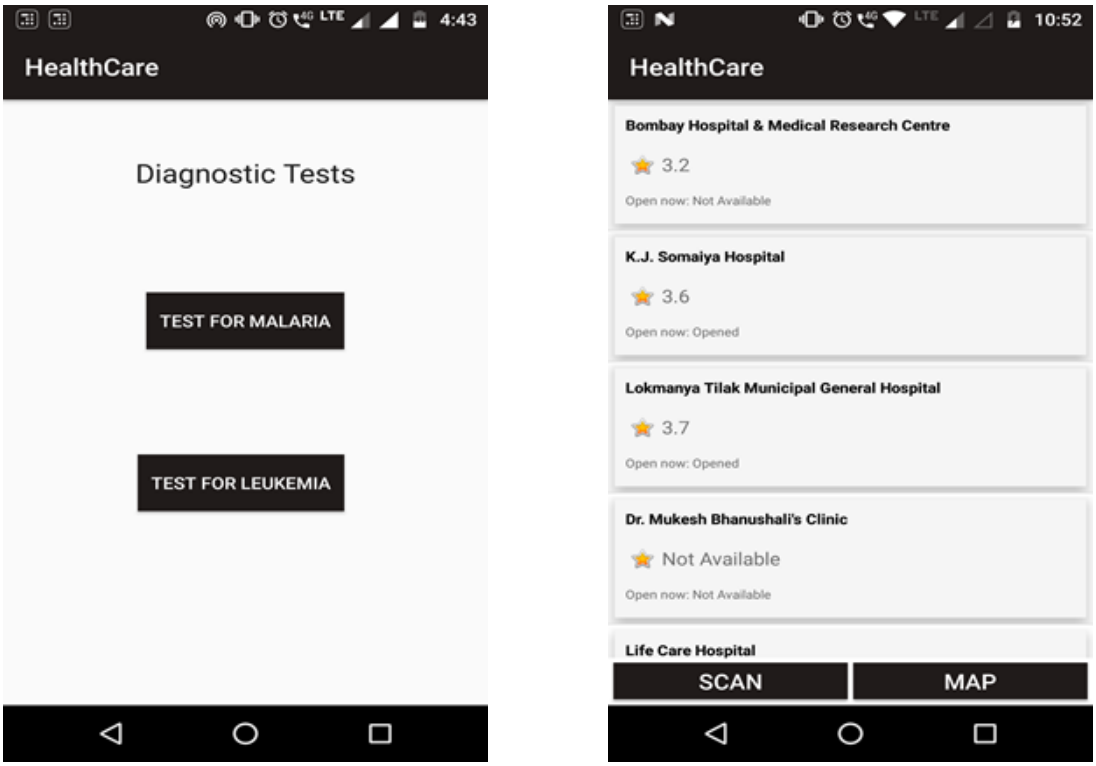
Python script to check for leukemia

```
import cv2
import numpy as np
import os
import pandas as pd

#img = cv2.imread('C:/Users/krishna/Downloads/Dataset/ALL_IDB1/im/
   Im024_1.jpg')



#testing
directory = 'C://xampp//htdocs//ImageUpload//uploads'
for file in os.listdir(directory):
   filename = os.fsdecode(file)
   if filename.endswith(".jpg"):
      img = directory+'//'+str(filename)
      img1 = cv2.imread(img)
      ORANGE_MIN = np.array([90, 120, 70],np.uint8)
      ORANGE_MAX = np.array([180, 255, 255],np.uint8)
      hsv_img = cv2.cvtColor(img1,cv2.COLOR_BGR2HSV)
      segment = cv2.inRange(hsv_img, ORANGE_MIN, ORANGE_MAX)
      imS = cv2.resize(segment, (960, 540))
      c = cv2.countNonZero(imS)
      if c>5386:
         print("Leukemia Positive")
      elsif c<3011:
         print("Leukemia Negative")
      else:
            print("At risk, consult a doctor")
      os.remove(img)
   else:
      continue
```

Chapter 6

Result/Analysis

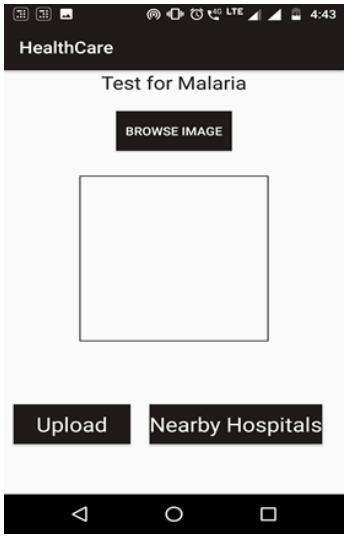
6.1 Screenshots of User Interface (UI)



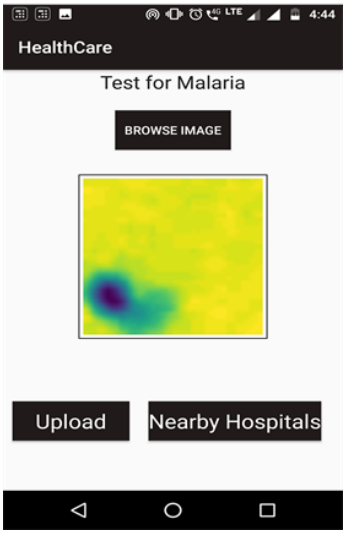
(a) Main Application View (b) Find Nearby Hospitals

Figure 6.1: Application View

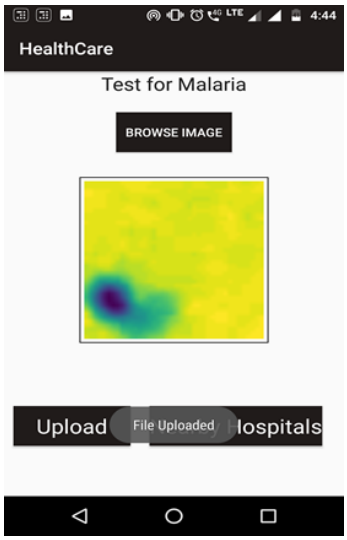
6.1.1 Malaria



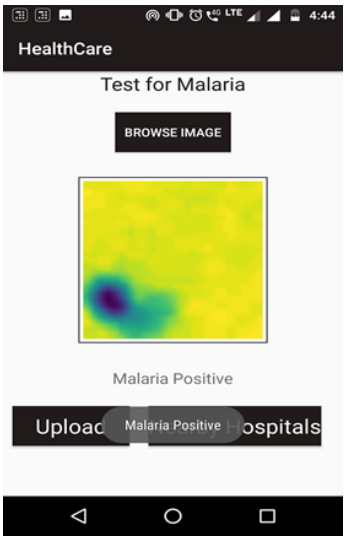
(a) App View for Malaria



(b) Test Image Selected (Malaria)



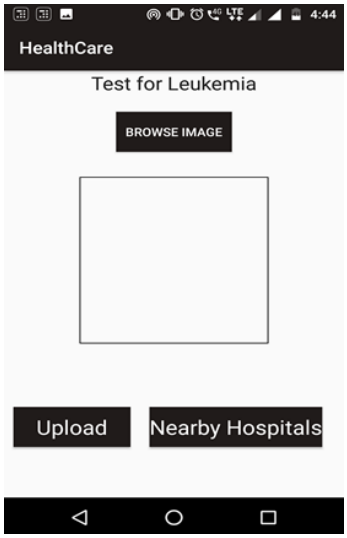
(c) Image Uploaded to server (Malaria)



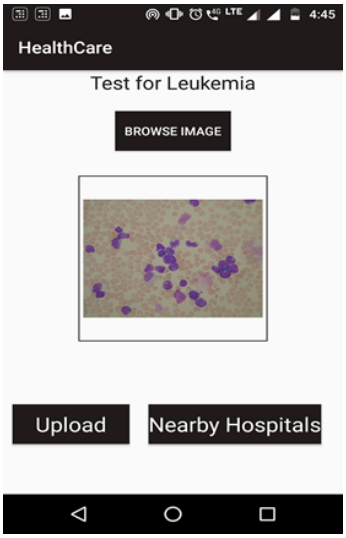
(d) Result returned from server (Malaria)

Figure 6.2: App View of Malarial Diagnosis

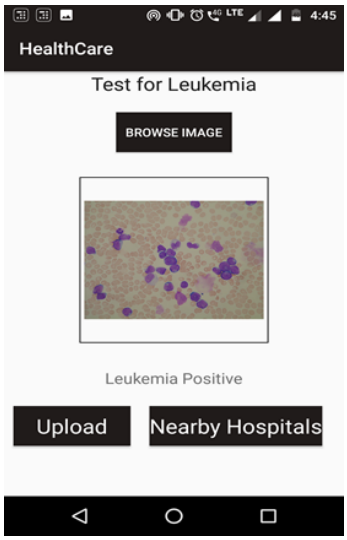
6.1.2 Leukemia



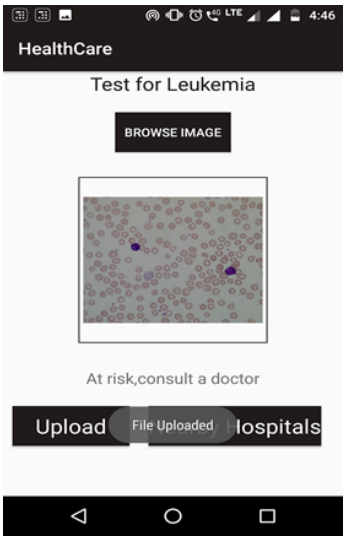
(a) App View for Leukemia



(b) Test Image Selected (Leukemia)



(c) Result positive (Leukemia)



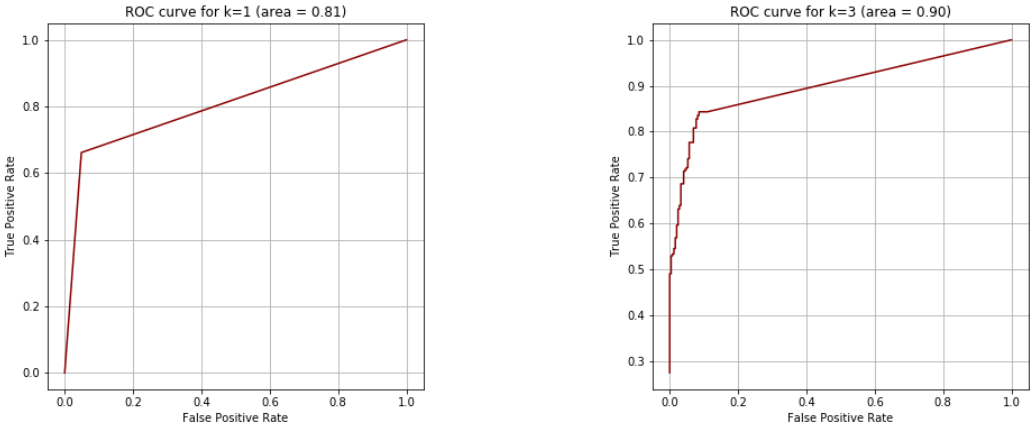
(d) Result doubtful (Leukemia)

Figure 6.3: App View of Leukemia Diagnosis

6.2 Graphs

6.2.1 KNN Classifier Accuracy

We use multiple values of k to generate our classifier and test for the best accuracy. We used receiver operator characteristics (ROC) curve for evaluation, with the area under the ROC curve (AUC) being the measure of how well a parameter distinguished between two diagnostic groups (diseased/normal). From our results, we select the classifier with highest area under curve, at k=5



(a) ROC Curve for k=1

(b) ROC Curve for k=3

Figure 6.4: ROC Curves for KNN Classifier

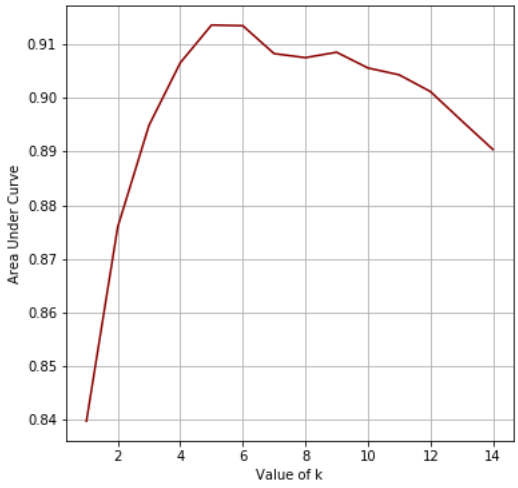


Figure 6.5: Curve plotting all values of AUC for different k

6.2.2 Leukemia Pixel Count

We attempt to find a range of values for which we can successfully diagnose sample as leukemia positive or negative.

Pixel Count	Negative Diagnosis	Positive Diagnosis
Min. Pixel Count	579	3011
Max. Pixel Count	5386	71237

Figure 6.6: Range of Pixel Counts for Leukemia Diagnosis

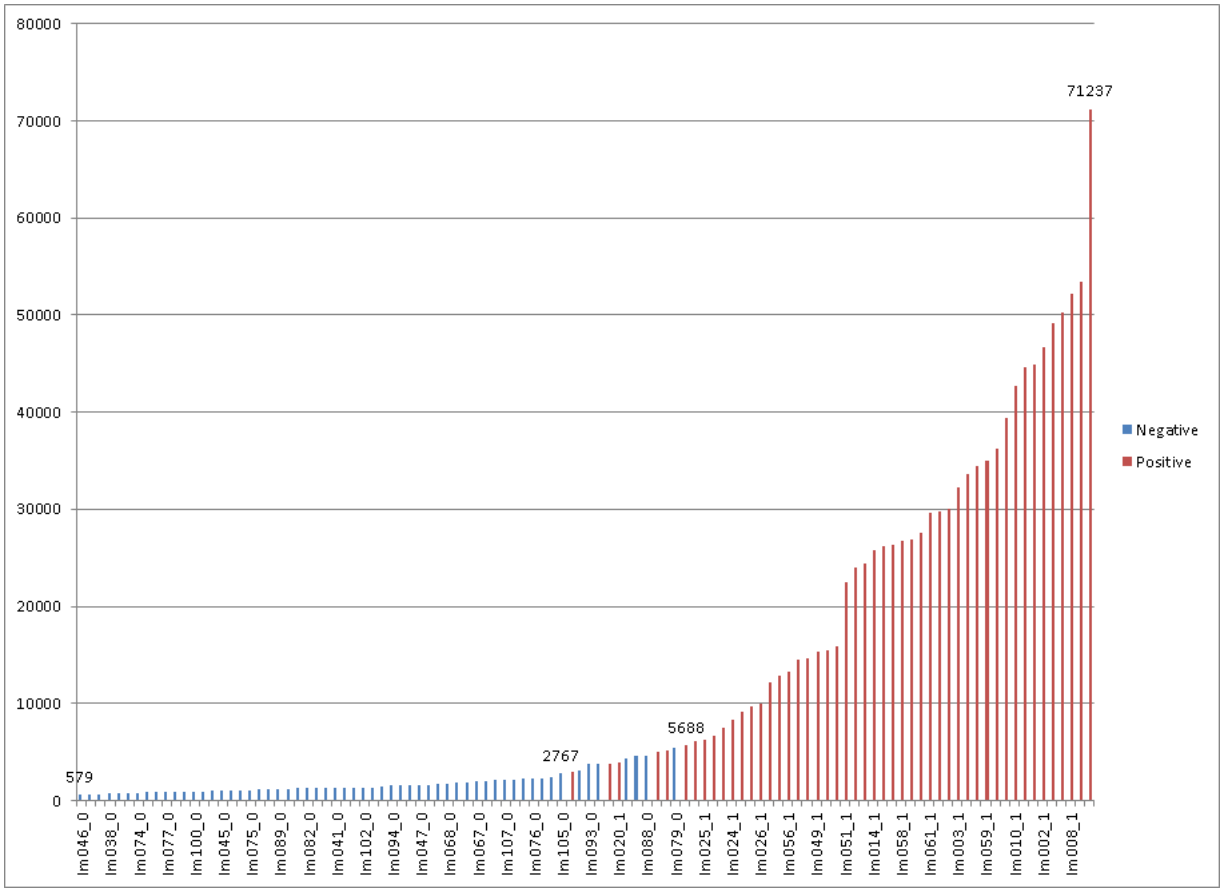


Figure 6.7: Pixel Counts on training dataset (Leukemia)

6.3 Gantt Chart

Task Name	Start Date	End Date	Status
Requirement Gatheri	08/21/17	09/26/17	Completed
Analysis	09/27/17	10/07/17	Completed
Design	10/08/17	10/31/17	Completed
Implementation	11/01/17	01/31/18	Completed
Testing	02/01/18	02/28/18	Completed
Deployment	03/01/18	03/19/18	Completed

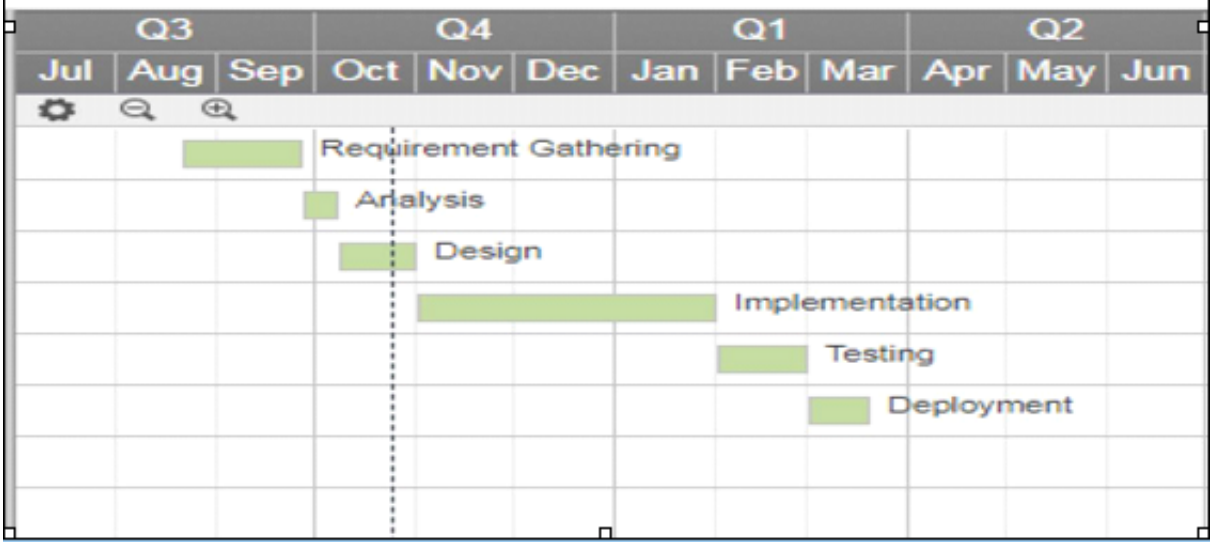


Figure 6.8: Gantt Chart Timeline for Project

Chapter 7

Conclusion

7.1 Limitations

1. **Scale of Image:** Dataset obtained may have images of varying sizes. Aligning or resizing such images is time consuming and redundant.
2. **Quality of image and stain:** Quality of images may affect system performance. Blur or low quality and unstained image if trained will generate wrong classifier which will further affect diagnosis.
3. **Obtaining dataset for training:** We need large datasets for training the classifier for high accuracy of the system. Getting proper datasets involves various barriers; also the obtained dataset may not be in the proper form or have acceptable quality.
4. **Hardware limitations:** The current system requires both server and smartphone for overall diagnosis.

7.2 Future Scope

1. **Independence:** Include entire functionality within the app to provide serverless and offline testing.
2. **Reliability:** Implement additional classifiers for increased accuracy.
3. **Multiple Diagnoses:** As the structure of diagnosis is the same, we can include various blood related diseases provided there is a sufficient amount of image dataset available

7.3 Conclusion

The system aims to restrict human error while detecting the presence of Malaria parasites and leukemia in the blood samples by using image acquisition, pre-processing & KNN classifier. Using KNN classifier we are getting accurate and required results in the short period of time. The system is robust and is unaffected by exceptional conditions and achieves specificity and positive or negative prediction values.

References

[1] http://www.who.int/malaria/media/world-malaria-day-2016/en/

[2] https://www.lls.org/llsorg.prod.acquia-sites.com/facts-and-statistics/facts-and-statistics-overview/facts-and-statistics

[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4062392/

Research Papers

[4] http://ieeexplore.ieee.org/document/7823258/

[5] http://ieeexplore.ieee.org/abstract/document/5712739/

[6] http://ieeexplore.ieee.org/document/6707610/

[7] http://ieeexplore.ieee.org/document/6179129/

[8] https://www.ncbi.nlm.nih.gov/pubmed/26409536

[9] http://ieeexplore.ieee.org/document/7867644/

[10] http://www.trp.org.in/wp-content/uploads/2016/12/AJEAT-Vol.5-No.2-July-Dec-2016-pp.1-5.pdf

[11] http://proceedings.mlr.press/v56/Quinn16.pdf

[12] http://ieeexplore.ieee.org/document/6949840/

[13] http://ijact.org/volume4issue1/IJ0410050.pdf

Datasets

[14] https://homes.di.unimi.it/scotti/all/-Acute

[15] http://www.biosigdata.com/-Malaria,Leishmaniasis