**CHAPTER 1**

**INTRODUCTION**

* 1. **Overview:**

Data mining is the process of finding previously unknown patterns and trends in databases and using that information to build predictive models. Data mining combines statistical analysis, machine learning and database technology to extract hidden patterns and relationships from large databases.

Diabetes is a situation which causes deficiency due to less amount of insulin in the blood. Warning sign of high blood sugar results in frequent urination, feeling thirsty, increased hunger. If it is not medicated, it will lead to many difficulties. This difficulty lead to death. Severe difficulties lead to cardiovascular disease foot sores, and eye blurriness. When there is a rise within the sugar level within the blood, it is referred to as prior diabetes. The prior diabetes isn't therefore great than the traditional worth. Diabetes is appreciations to either the exocrine gland not manufacturing plentiful hypoglycemic agent not responding properly to the hypoglycemic agent created.

The World Health Statistics 2012 report enlightens the fact that one in three adults worldwide has raised blood pressure - a condition that causes around half of all deaths from stroke and heart disease. Heart disease, also known as cardiovascular disease (CVD), encloses a number of conditions that influence the heart – not just heart attacks. Heart disease was the major cause of casualties in the different countries including India. Heart disease kills one person every 34 seconds in the United States.

* 1. Introduction:

In recent years, the application of machine learning techniques in healthcare has garnered significant attention due to its potential to revolutionize disease diagnosis, prognosis, and treatment. One particularly promising area within this field is the prediction of multiple diseases simultaneously using machine learning algorithms. Traditional medical diagnosis often relies on the expertise of healthcare professionals and may involve extensive manual analysis of patient data, which can be time-consuming and prone to errors [1].

Additionally, healthcare resources are often limited, and early detection of multiple diseases can be challenging, leading to delayed treatment and poorer outcomes for patients. Machine learning offers a solution to these challenges by enabling the automated analysis of large volumes of diverse healthcare data, including medical records, imaging studies, genetic information, and more. By

training machine learning models on such data, it becomes possible to develop predictive algorithms capable of identifying patterns and associations indicative of various diseases [3].

Multiple disease prediction using machine learning involves the development of algorithms that can simultaneously predict the likelihood of several different diseases based on a patient's clinical data. These algorithms can leverage a wide range of input features, including demographic information, medical history, laboratory test results, imaging findings, lifestyle factors, and genetic markers.

**1.3 Motivation:**

**1. Improved Patient Health Outcomes:** One of the primary motivations for employing machine learning for the prediction of multiple diseases, including diabetes, heart disease, and leukemia, is to improve patient health outcomes.

**2. Personalized Medicine:** Machine learning-based disease prediction facilitates the delivery of personalized medicine by tailoring interventions to individual patient characteristics. This motivation aims to optimize treatment efficacy, minimize adverse effects, and enhance patient satisfaction through individualized care approaches.

**3. Preventive Healthcare:** Predictive modeling enables proactive healthcare management by identifying individuals at high risk of developing specific diseases before the onset of clinical symptoms. This motivation aims to shift healthcare focus from reactive treatment to proactive disease prevention, ultimately reducing healthcare costs and improving population health outcomes.

**4. Resource Optimization:** Another motivation for disease prediction using machine learning is the efficient allocation of healthcare resources. This motivation aims to address resource constraints, enhance healthcare system sustainability, and improve access to quality care for all individuals.

**5. Clinical Decision Support:** Predictive models integrated into clinical decision support systems (CDSS) serve as valuable tools for healthcare providers, offering evidence-based insights and recommendations to inform diagnostic and treatment decisions. This motivation aims to empower healthcare providers with actionable information derived from machine learning algorithms, thereby improving clinical decision-making processes and enhancing patient care quality.

**1.4 Objectives:**

1. **Early Detection:** The primary objective of utilizing machine learning for the prediction of multiple diseases, including diabetes, heart disease, and
2. leukemia, is to achieve early detection. By leveraging predictive models trained on comprehensive datasets, the aim is to identify individuals at risk of developing these conditions before the onset of clinical symptoms.
3. **Precision Medicine:** Another objective is to advance the paradigm of precision medicine by tailoring interventions and treatment plans to the individual characteristics of each patient. Machine learning models can analyze diverse data sources, such as genetic profiles, clinical records, and lifestyle factors, to create personalized risk profiles.
4. **Risk Stratification:** Machine learning-based prediction models aim to stratify individuals into different risk categories based on their likelihood of developing specific diseases. By quantifying risk levels, healthcare providers can prioritize preventive interventions and allocate resources more efficiently.
5. **Clinical Decision Support:** By providing clinicians with real-time risk assessments and predictive analytics, CDSS can enhance diagnostic accuracy, facilitate personalized treatment planning, and improve patient management strategies. This objective aims to empower healthcare providers with actionable insights derived from machine learning algorithms, thereby enhancing the quality and efficiency of healthcare delivery.
6. **Research and Innovation:** The objective of employing machine learning for disease prediction is to foster ongoing research and innovation in the field of healthcare This objective aims to drive progress in understanding disease mechanisms, discovering new therapeutic targets, and ultimately improving patient outcomes through innovative healthcare solutions.
7. **Ethical and Regulatory Compliance:** Ensuring ethical and regulatory compliance is a fundamental objective of utilizing machine learning for disease prediction. This objective aims to establish robust governance frameworks, adhere to legal and regulatory requirements, and mitigate potential risks associated with data misuse or algorithmic bias, thereby fostering trust and confidence in machine learning-based healthcare applications.

**1.5 Organizations of Rest of Report:**

Rest of the report is organized as follows:

1. Chapter one introduces the project, providing a brief overview of Multiple Disease Prediction By Machine Learning.
2. Chapter two comprises a literature survey summarizing all the referenced research papers.
3. Chapter 3 includes all technologies, tools used, requirements, and some methodology employed in the project.
4. In the fourth chapter, System Implementation and Testing of Multiple Disease Prediction By Machine Learning are discussed using Architecture, Techniques, and various Algorithms.
5. The fifth Chapter presents the Results of Multiple Disease Prediction By Machine Learning, including all working output images.
6. The Sixth chapter serves as the Conclusion of the Overall Project.
7. The last chapter focuses on the Future Scope of Multiple Disease Prediction By Machine Learning.

**CHAPTER 2**

**LITERATURE REVIEW**

In this chapter, theb review of research paper referred and used in this project are discussed.

Sai and Reddy (2022) proposed a predictive model for “heart disease utilizing the Artificial Neural Network (ANN) algorithm within the domain of data mining”. This work addresses the increasing costs associated with heart disease diagnosis, highlighting the necessity for novel systems capable of accurate disease prediction. The proposed model evaluates patient conditions based on diverse parameters including heart rate, blood pressure, cholesterol levels, and more [1].

Ghane, S., Bhorade, N., Chitre, N., Poyekar, B., Mote, R., & Topale, P. (2021). "Diabetes Prediction using Feature Extraction and Machine Learning Models." In this study, the authors propose machine learning algorithms, namely KNN, SVM, Decision Tree, Random Forest, LGBM, and Adaboost, for predicting diabetes. [2].

Chadha, G. K., Srivastava, A., Singh, A., Gupta, R., & Singla, D. (2022). "An Automated Method for Counting Red Blood Cells using Image Processing." In healthcare, blood testing is regarded as one of the most significant medical examinations. Pathology labs often count different types of blood cells to diagnose various diseases in patients. Counting Red Blood Cells (RBCs) in images of blood cells can play a crucial role in detecting and monitoring the treatment progress of numerous diseases, such as Acute Lymphoblastic Leukemia [3].

(Science & Faculty, 2023) suggested "Heart Disease Prediction using Data Mining and Machine Learning Algorithms." This study aims to extract hidden patterns through the application of data mining techniques. The study found that the J48 algorithm, based on UCI data, demonstrated higher accuracy rates compared to LMT [4].

Hasan, M. K., Alam, M. A., Das, D., Hossain, E., & Hasan, M. (2022). "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers." This study proposes a robust framework for diabetes prediction. The framework includes outlier rejection, filling missing values, data standardization, feature selection, K-fold cross-validation, and the utilization of various Machine Learning (ML) classifiers such as k-nearest Neighbor, Decision Trees, Random Forest, AdaBoost, Naive Bayes, XGBoost, and Multilayer Perceptron (MLP). Additionally, the study introduces weighted ensembling of different ML models

to enhance diabetes prediction, where the weights are estimated from the corresponding Area Under ROC Curve (AUC) of the ML model [5].

George, J., Sobha, T., Ashok V., A., & Eldhose, J. (2020). “An efficient image processing technique to count red blood cells”. In this study Shortage of red blood cell (RBC), that constitutes 99 percent of blood cells and specialized as oxygen carrier, causes various blood disorders. The RBC is the important parameter while diagnosis and pathological study [6].

A, A. S., & Naik. (2021). "Diagnosis of Heart Disease from Patient’s Medical Dataset." This project focuses on developing a prediction system for diagnosing heart disease using patient medical datasets. The study employs k-means and Naive Bayes algorithms for heart disease prediction. The project utilizes data mining techniques such as clustering and classification methods to extract knowledge from the Cleveland Heart Database, consisting of 13 attributes with a total of 300 records [7].

Beyene, A., & Kamat, S. (2021). "Heart Disease Prediction System using Data Mining Techniques." This project utilizes various algorithms including Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM, and ANN for heart disease prediction. The study recommends the use of SVM for its effectiveness and higher accuracy compared to other data mining algorithms. The WEKA software is employed for automatic disease diagnosis and improving healthcare services [8].

Rahadi, Hoodooing, M. C., & Hoodooing, A. (2022). "Red Blood Cells and White Blood Cells Detection by Image Processing. This project aims to develop a computer program for automated detection and identification of red and white blood cells using image processing techniques [9].

Sonar, P., & JayaMalini, K. (2020). "Diabetes Prediction Using Different Machine Learning Approaches." This study focuses on developing a system for predicting the risk level of diabetes in patients. The model employs Decision Tree, ANN, Naive Bayes, and SVM algorithms, achieving significant accuracy rates [10].

Sultana, S., Haider, S. S., & Uddin, M. S. (2022). "Analysis of Cardiovascular Disease." This project proposes data mining techniques for predicting cardiovascular diseases, aiming to provide a survey of current techniques to extract information from datasets. The study focuses on predicting diseases with a smaller number of attributes [11].

Soni, A., Ansari, M. A., & Sharma, M. (2021). "Non-linear Classification Algorithm for Heart Disease Prediction." This study investigates the use of different data mining techniques for predicting heart diseases, utilizing big data tools such as Hadoop Distributed File System (HDFS) and parallel SVM for optimized attribute set prediction [12].

Bhavana, R., Srivastava, P., N, S., & Reddy, R. (2021). "Review on Identification of Red Blood Cells by Image Processing." This project focuses on automatic diagnosis of microscopic blood smear images by identifying and categorizing red blood cells using image processing techniques [13].

Shetty, D., Rit, K., Shaikh, S., & Patil, N. (2023). "Diabetes Disease Prediction Using Data Mining." This study aims to develop an Intelligent Diabetes Disease Prediction System utilizing Bayesian and KNN algorithms for analyzing diabetes patient databases [14].

Kirmani, A. (2021). "Multi-disease Prediction Using Data Mining Techniques." This project utilizes data mining techniques for predicting multiple diseases including heart disease, diabetes, and breast cancer, aiming to reduce the number of tests required for prediction [15].

In this way we summarized the research papers we refer for the project development and report preparation. These papers helps in developing the correct Disease Prediction module with high precision and accuracy. All the papers mention above have significant importance throughout the development of the project.

**CHAPTER 3**

**SYSTEM DESIGN**

In our project the hardware and software which we have used are as follows:

**3.1 Hardware Requirements**

* + **Server Infrastructure:** High-performance server hardware to host the centralized database and application logic. Redundancy and backup systems to ensure data integrity and availability.
  + **User Devices:** Desktop computers, laptops, or tablets for administrators and caregivers to access the An Automatic Detection & Counting of RBCs using Image Processing.
  + **Processor:** Minimum Intel Pentium IV or Above
  + **RAM:** 4 GB or above.
  + **Storage:** 50 GB or above

**3.2 Software Requirements**

* + **Operating System:** Minimum Windows 10 or above.
  + **Browser:** Chrome, Edge, Brave or any type of browser.
  + **User Interface (UI) Technologies:** HTML, CSS, and JavaScript for building the user interface. Consider responsive design principles for accessibility on various devices.
  + **Programming Languages:** Used programming languages suitable for web application development, such as Python 3.12, Django Framework, SQLite Database.
  + **IDE:** VSCode (Integrated Development Environment).

**3.3 Frontend Technologies**

* + **HTML:** HTML stands for Hypertext Mark-up Language. It is a mark-up language used to create and design web pages. HTML provides a way for developers to structure content on a webpage using a set of tags and attributes that define how the content is displayed.
  + **CSS:** CSS stands for Cascading Style Sheets. It is a stylesheet language used for describing the presentation of web pages, including their layout, colors, fonts, and other visual aspects. CSS provides a way for developers to separate the content of a web page from its presentation, making it easier to maintain and update the page. With CSS, developers can create a set of rules that apply to all pages on a website, ensuring a consistent look and feel throughout the site.
  + **JavaScript:** JavaScript is a programming language used to create interactive and dynamic web pages. It is an essential component of modern web development, used for everything from simple form validation to building complex web applications. JavaScript is a client-side language, meaning that it is executed in the user’s web browser rather than on the server.

**3.4 Backend Technologies**

* + **Python Language:** Python is a high-level, interpreted, and general-purpose programming language. Python was created by Guido van Rossum and was first released in 1991. Python is an interpreted language, which means that the Python code is executed line by line by an interpreter, rather than being compiled into machine code
  + **Django Framework:** Django is a high-level Python web framework that enables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It is free and open source, has a thriving and active community, great documentation, and many options for free and paid-for support.
  + **SQLite Database:** SQLite is a self-contained, serverless, and zero-configuration relational database management system (RDBMS). Here are some key points about SQLite.

**3.5 Methodology**

The process comprises of image acquisition followed by numerous image processing operations which are segmentation, median filtering, edge detection, erosion and counting of cells. The counting of Red blood cells will be done on the basis of edge detection using Hough transformation. While developing this project we used SVM is supervised learning algorithm used.

**3.5.1 Image Acquisition**

* + **Source of Blood Smear Images:** Collect high-resolution digital images of blood smears using a suitable microscope equipped with a digital camera. Ensure to diverse samples to account for variations in staining, magnification, and lighting conditions.
* **Image Preprocessing:** Apply preprocessing techniques to enhance image quality and reduce noise. This includes normalization, contrast adjustment, and noise reduction to standardize the images for further analysis.

**3.5.2 Image Preprocessing & Enhancement**

* **Blood Smear Images:** Collect high-resolution digital images of blood smears using a suitable microscope equipped with a digital camera. Ensure diverse samples to account for variations in staining, magnification, and lighting conditions.
* **Image Enhancement Techniques:** Apply preprocessing techniques to enhance image quality and reduce noise. This includes normalization, contrast adjustment, and noise reduction to standardize the images for further analysis.

**3.5.3 Image Segmentation**

The segmentation phase, which is concerned with extracting individual object components carrying pivotal information. Image segmentation can also be performed under morphological operations since in medical image segmentation, morphology plays an important role.

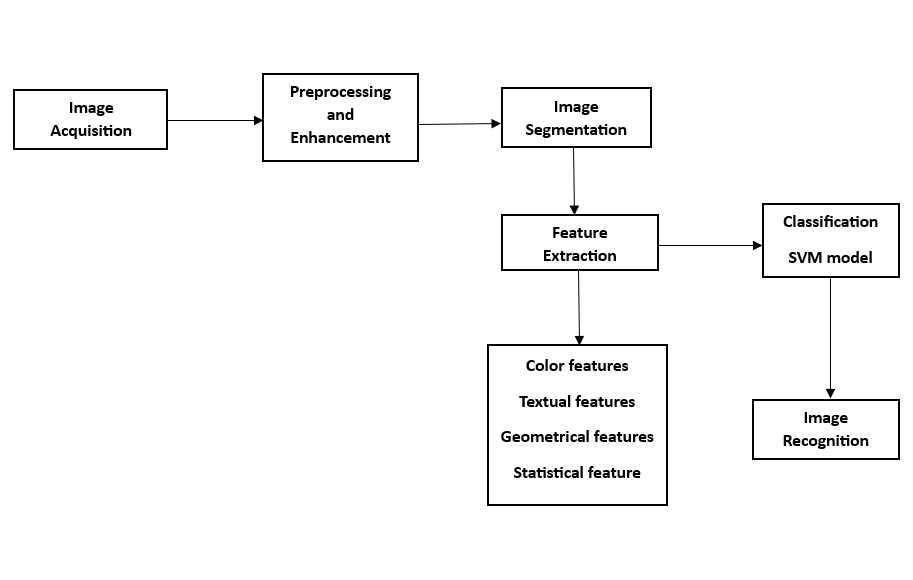
* + **Thresholding:** Apply adaptive thresholding techniques to separate RBCs from the background. Experiment with various thresholding methods to find the most effective approach for different images.
  + **Morphological Operations:** Morphological operations apply a structuring element to an input image, creating an output image of the same size. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries.
  + **Feature Extraction:** In feature extraction we extract relative information from the images. We extract the features such as Color, Geometry, Texture features and statistical features. The goal of feature extraction is to obtain a set of image descriptors. By finding the relationship between
  + these descriptors, the patterns determining the images can be discovered. The accurate feature extraction and leukemia classification are proportionately dependent on the correct segmentation of the maximized and cropped lymphocytes. Feature extraction is a part of the dimensionality reduction process, in which an initial set of the raw data is divided and reduced to more manageable groups. So when you want to process it will be easier.
  + **Image Classification:** A Support Vector Machine (SVM) is a binary linear classification whose decision boundary is explicitly constructed to minimize generalization error. It is a very powerful and versatile Machine Learning model, capable of performing linear or nonlinear classification, regression and even outlier detection.

**CHAPTER 4**

**SYSTEM IMPLEMENTATION AND TESTING**

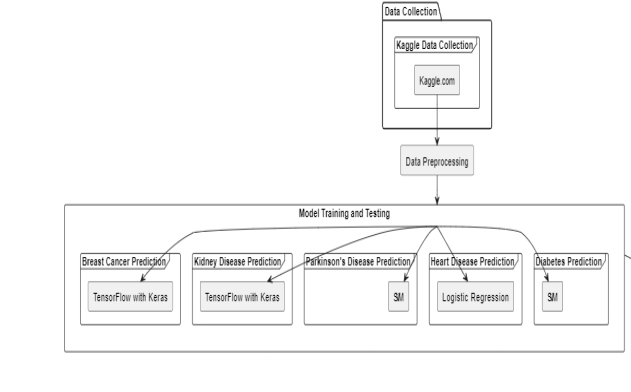
In this chapter, the system implementation and testing is discussed in details

**4.1 Proposed Architecture**

The system architecture of the project is as follows.

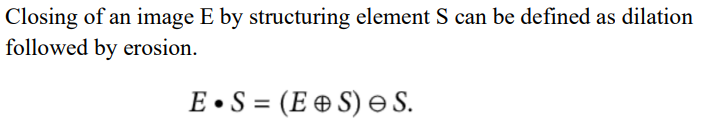
**Figure 4.1:** Proposed Architecture

**4.2 Data Collection**



**Figure 4.2:** Data Collection

**4.3 Various Techniques :**

1. **Data Collection and Preprocessing:** Gather datasets containing relevant features and labels for each disease. This data could come from various sources such as medical records, research studies, or publicly available datasets. Preprocess the data to handle missing values, outliers, and inconsistencies. This might involve techniques like imputation, normalization, and feature scaling. Split the data into training and testing sets to evaluate the model's performance.
2. **Feature Selection and Engineering:** Identify important features that are predictive of each disease. This can be done through statistical analysis, domain knowledge, or feature selection algorithms. Perform feature engineering to create new features or transform existing ones to improve the model's predictive power. This could include techniques like dimensionality reduction, polynomial features, or encoding categorical variables.
3. **Model Selection and Training:** Choose appropriate machine learning algorithms for each disease prediction task. Common algorithms include logistic regression, decision trees, random forests, support vector machines, and neural networks. Train separate models for each disease using the training data. Tune hyperparameters using techniques like cross-validation to optimize model performance. Consider ensemble methods to combine multiple models for improved accuracy and robustness.
4. **Evaluation and Validation:** Evaluate the trained models using the testing dataset. Use evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC) to assess performance. Validate the models on independent datasets if available to ensure generalizability. Perform error analysis to understand where the models are making mistakes and iteratively refine them.
5. **Deployment and Integration:** Deploy the trained models into a production environment where they can be used to make predictions on new data. Integrate the models into a user-friendly interface such as a web application or API, allowing healthcare professionals or patients to input their data and receive disease predictions. Implement security and privacy measures to protect sensitive medical information.
6. **Monitoring and Maintenance:** Continuously monitor the performance of the deployed models in real-world settings. Retrain the models periodically using updated data to account for changes in disease patterns or patient demographics. Address any issues or feedback from users to ensure the system remains accurate and reliable over time.
7. **Image Acquisition:** Use a high-resolution microscope or a digital camera to capture images of blood samples containing RBCs. Ensure proper lighting conditions and color calibration for accurate image processing.
8. **Image Segmentation:** The segmentation phase, which is concerned with extracting individual object components carrying pivotal information. Image segmentation can also be performed under morphological operations since in medical image segmentation, morphology plays an important role. The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image. In the morphological dilation and erosion operations, the state of any given pixel in the output image is determined by applying a rule to the corresponding pixel and its neighbors in the input image. After performing the morphological operation, we will be able to get our region of and hence it undergoes feature extraction.
9. **Image Recognition:** After the classification model we can predict if the image is suffering from which type of leukemia like Acute lymphocytic leukemia (ALL), Acute myelogenous leukemia (AML). The final recognition of ALL from peripheral blood smear images is accomplished by optimized support vector machine (SVM). The SVM model predicts the disease and if it is suffering then it will try to make a circle in the image and hence it is recognized. The proposed model is comparatively effective and reliable.

**4.4 Algorithms:**

1. **Preprocessing**

ab\_global=[]

for image in images:

file\_name=dirname.file

input\_image=read from the directory

inputImLAB = rgb2lab(input\_image);

% Extract a\* and b\* channels and reshape

ab = double(inputImLAB(:,:,2:3));

ab = reshape(ab,nrows\*ncols,2);

ab\_global = [ab\_global;ab];

enhanced=contrast\_stretching(preprocessed\_image)

Erosion=erosion(enhanced)

Dilation=dilation()

segmented\_image = imbinarize(min\_data\_3d\_ind)

1. **Feature Extraction**

Compute radius, area, diameter, convexity…

centers = reshape([stats.Centroid], 2, CC.NumObjects);

majors = [stats.MajorAxisLength];

minors = [stats.MinorAxisLength];

diameters= [stats.EquivDiameter];

radii = [diameters./2];

**Load into the dataset**

stats(i).Contrast = out.Contrast;

stats(i).Correlation = out.Correlation;

stats(i).Energy = out.Energy;

stats(i).Homogeneity = out.Homogeneity;

stats(i).ent = ent\_result;

stats(i).std\_dev =std\_dev ;

stats(i).skew = skew ;

stats(i).meanval = meanval;

**SVM (Support Vector Machine)**

Break dataset into training and testing

model = svm(train\_data,test\_data)

prediction=model(dependencies)

print(df.head())

#To delete small regions...

df = df[df['area'] > 50]

print(df.head())

#Convert to micron scale

df['area\_sq\_microns'] = df['area'] \* (pixels\_to\_um\*\*2)

df['equivalent\_diameter\_microns'] = df['equivalent\_diameter'] \* (pixels\_to\_um)

print(df.head()) df.to\_csv('safal.csv')

def featureExtraction(img):

cells=img[:,:,0]

pixels\_to\_um = 0.454

ret1, thresh = cv2.threshold(cells, 0, 255,

cv2.THRESH\_BINARY+cv2.THRESH\_OTSU)

kernel = np.ones((3,3),np.uint8)

opening = cv2.morphologyEx(thresh,cv2.MORPH\_OPEN,kernel, iterations = 2)

from skimage.segmentation import clear\_border

opening = clear\_border(opening)

plt.imshow(opening, cmap='gray')

#This is our image to be segmented further using watershed

sure\_bg = cv2.dilate(opening,kernel,iterations=10)

plt.imshow(sure\_bg, cmap='gray') #Dark region is our sure background

dist\_transform = cv2.distanceTransform(opening,cv2.DIST\_L2,5)

plt.imshow(dist\_transform, cmap='gray') #Dist transformed img.

print(dist\_transform.max()) #gives about 21.9

ret2, sure\_fg = cv2.threshold(dist\_transform,0.5\*dist\_transform.max(),255,0)

plt.imshow(sure\_fg, cmap='gray')

sure\_fg = np.uint8(sure\_fg) #Convert to uint8 from float

unknown = cv2.subtract(sure\_bg,sure\_fg)

plt.imshow(unknown, cmap='gray')

ret3, markers = cv2.connectedComponents(sure\_fg)

plt.imshow(markers)

markers = markers+10

# Now, mark the region of unknown with zero

markers[unknown==255] = 0

plt.imshow(markers, cmap='jet') #Look at the 3 distinct regions.

#Now we are ready for watershed filling.

markers = cv2.watershed(img,markers)

props = measure.regionprops\_table(markers, cells,

properties=['label',

'area', 'equivalent\_diameter',

'mean\_intensity', 'solidity', 'orientation',

'perimeter'])

1. **Classification / Recognition (SVM)**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns from wrapper

import leukemia

%matplotlib inline cancer=read\_csv()

df\_cancer = pd.DataFrame(np.c\_[cancer['data'], cancer['target']], columns = np.append(cancer['feature\_names'], ['target']))

df\_cancer.head()

X = df\_cancer.drop(['target'], axis = 1)

# We drop our "target" feature and use all the remaining features in our dataframe to train the model.

X.head()

from sklearn.model\_selection

import train\_test\_split

y = df\_cancer['target']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.2, random\_state = 20)

from sklearn.svm import SVC svc\_model = SVC()

svc\_model.fit(X\_train, y\_train) y\_predict = svc\_model.predict(X\_test)

from sklearn.metrics

import classification\_report, confusion\_matrix

dataread=pd.read\_csv('safal.csv')

cm = np.array(confusion\_matrix(y\_test, y\_predict, labels=[1,0]))

confusion = pd.DataFrame(cm, index=['is\_cancer', 'is\_healthy'], columns=['ALL','MLL'])

print(string1)

1. **Histogram**

def histeq(img):

a = np.zeros((256,),dtype=np.float16)

b = np.zeros((256,),dtype=np.float16)

imghist = img

height,width=img.shape

#finding histogram

for i in range(width):

for j in range(height):

g = imghist[j,i]

a[g] = a[g]+1

#performing histogram equalization

tmp = 1.0/(height\*width)

b = np.zeros((256,),dtype=np.float16)

for i in range(256):

for j in range(i+1):

b[i] += a[j] \* tmp

b[i] = round(b[i] \* 255)

# b now contains the equalized histogram

b=b.astype(np.uint8)

return imghist

1. **Edge Detection**

def edgeDetection(img):

imgS = img.astype(np.float16)

sobx=[[-1, -2, -1],

[0, 0, 0],

[1, 2, 1]]

sobx = np.array(sobx, np.float16)

soby =[[-1, 0, 1],

[-2, 0, 2],

[-1, 0, 1]]

soby = np.array(soby,np.float16)

for i in range(1,254):

for j in range(1,254):

imgtemp = img[i-1:i+2, j-1:j+2]

x = np.sum(np.multiply(sobx,imgtemp))

y = np.sum(np.multiply(soby,imgtemp))

pixvalue = np.sqrt(x\*\*2 + y\*\*2)

imgS[i,j] = pixvalue img

S = imgS.astype(np.uint8)

return imgS

1. **Detect Circles**

def detectCircles(img,openedimg):

imgcircle = cv2.cvtColor(img, cv2.COLOR\_GRAY2BGR)

detected\_circles = cv2.HoughCircles(openedimg,

cv2.HOUGH\_GRADIENT, 10, minDist= 10, param2= 30, minRadius = 1,

maxRadius = 13)

ctr=0

if detected\_circles is not None:

# Convert the circle parameters a, b and r to integers.

detected\_circles = np.uint16(np.around(detected\_circles))

for pt in detected\_circles[0, :]:

a, b, r = pt[0], pt[1], pt[2] #a,b are the coord

imgcirclefinal = cv2.circle(imgcircle, (a, b), r, (0, 255, 0), 2)

# Draw a small circle (of radius 1) to show the center.

#cv2.circle(img1, (a, b), 1, (255, 0, 0), 3)

ctr+=1

return imgcirclefinal, ctr

1. **Contract Streching**

def contrastStretching(img, r1, r2,a,b,c):

s1 = a\*r1

s2 = b\*(r2-r1)+s1

imgC = np.zeros((256,256), dtype=np.int32)

for i in range(0,256):

for j in range(0,256):

r = img[i,j]

if r<r1:

imgC[i,j] = a\*r

elif r>r1 and r<r2:

imgC[i,j] = b\*(r-r1) +s1

else:

imgC[i,j] = c\*(r-r2) + s2

imgC = imgC.astype(np.uint8)

return imgC

1. **Erosion and Dilation**

def dilation(img,mask):

img = img.astype(np.float16)

dilimg = np.zeros((256,256), dtype=np.float16)

for i in range(1,255):

for j in range(1,255):

imgtemp = img[i-1:i+2, j-1:j+2]

res = np.multiply(imgtemp,mask)

dilimg[i,j] = np.amax(res)

dilimg = dilimg.astype(np.uint8)

return dilimg

def erosion(img,mask):

img = img.astype(np.float16)

eroimg = np.zeros((256,256), dtype=np.float16)

for i in range(1,255):

for j in range(1,255):

imgtemp = img[i-1:i+2, j-1:j+2]

res=[]

for k in range(0,3):

for m in range(0,3):

if mask[k][m] ==1:

a = imgtemp[k,m]

res.append(a)

eroimg[i,j] = np.amin(res)

eroimg = eroimg.astype(np.uint8)

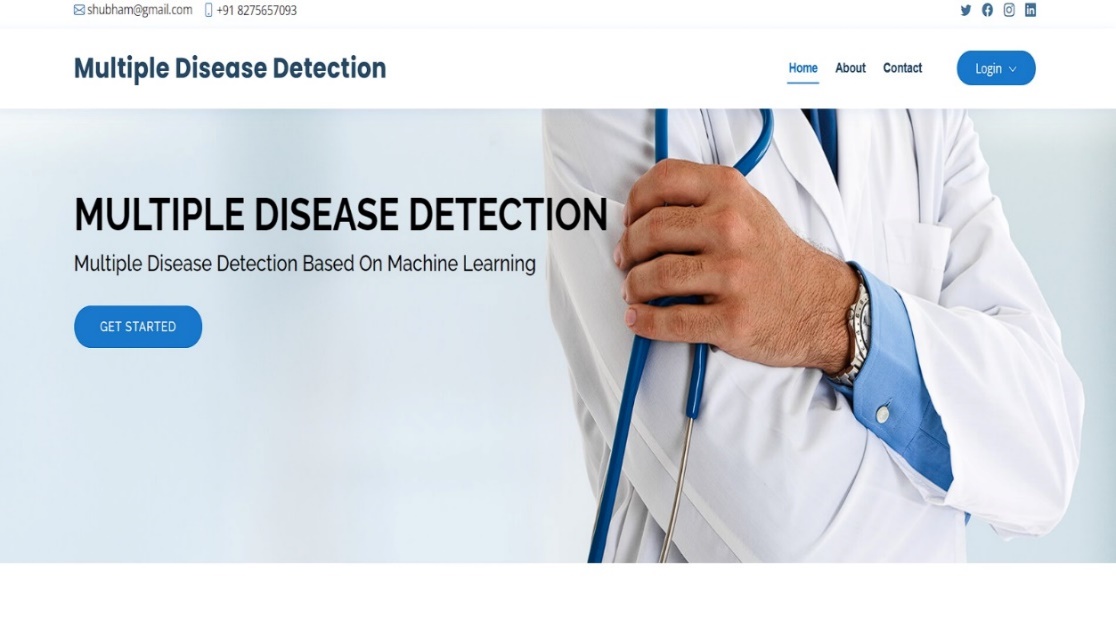
return eroimg

**CHAPTER 5**

**RESULT AND DISCUSSION**

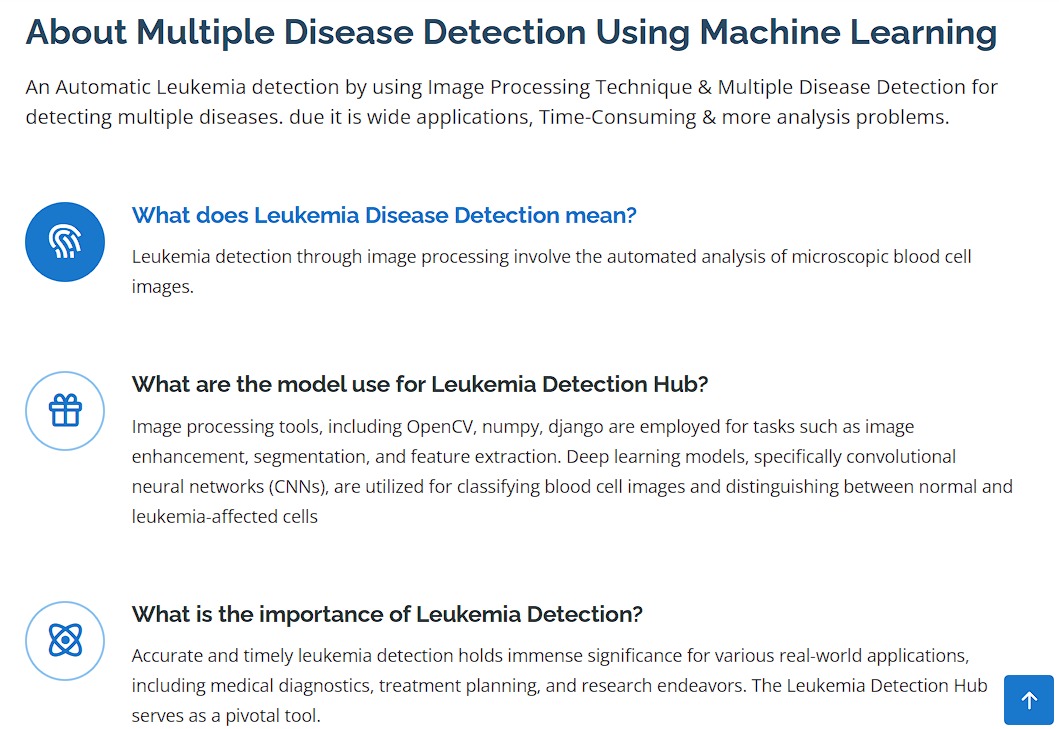
This is what you see when you go to the web interface. You are supposed to copy the news and paste it into the input box.

**5.1 Home Page:** It is our starting page which is open after application has started.

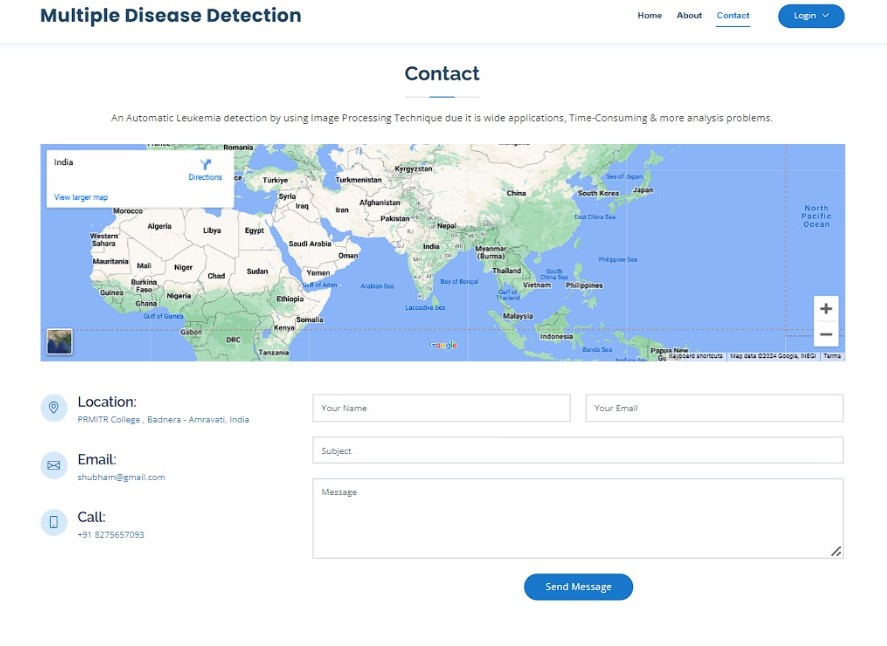


**Screenshot 5.1: Home Page**

**5.2 About Page:** This is page that gives the information about leukemia disease and all other disease.

****

**Screenshot 5.2: About Page**

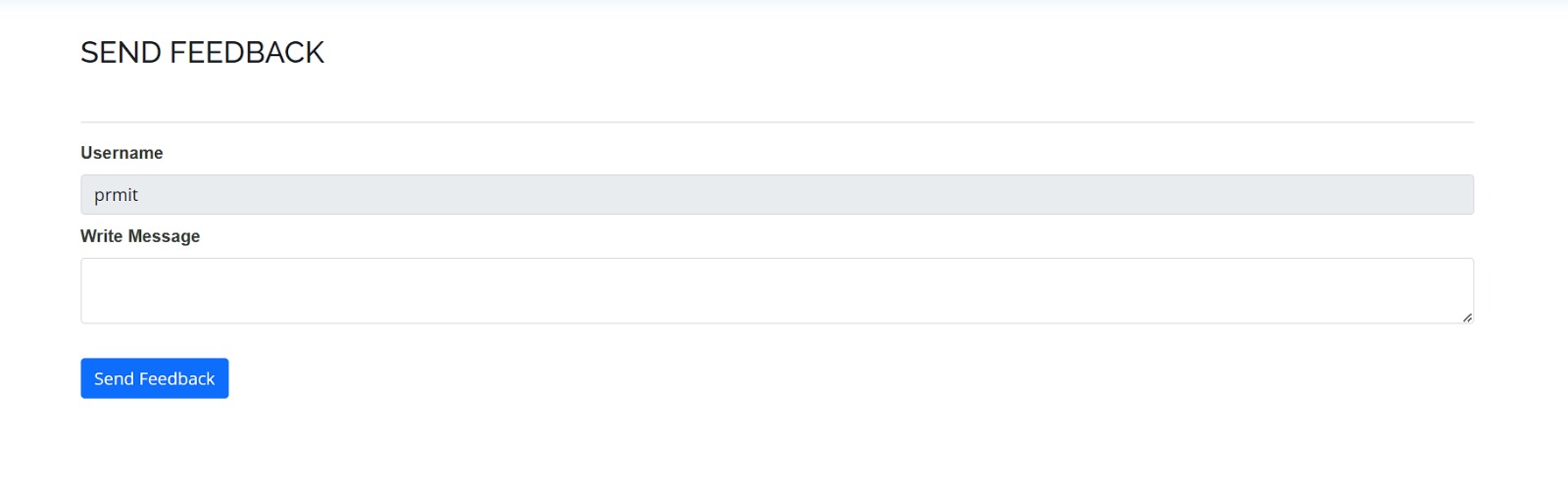
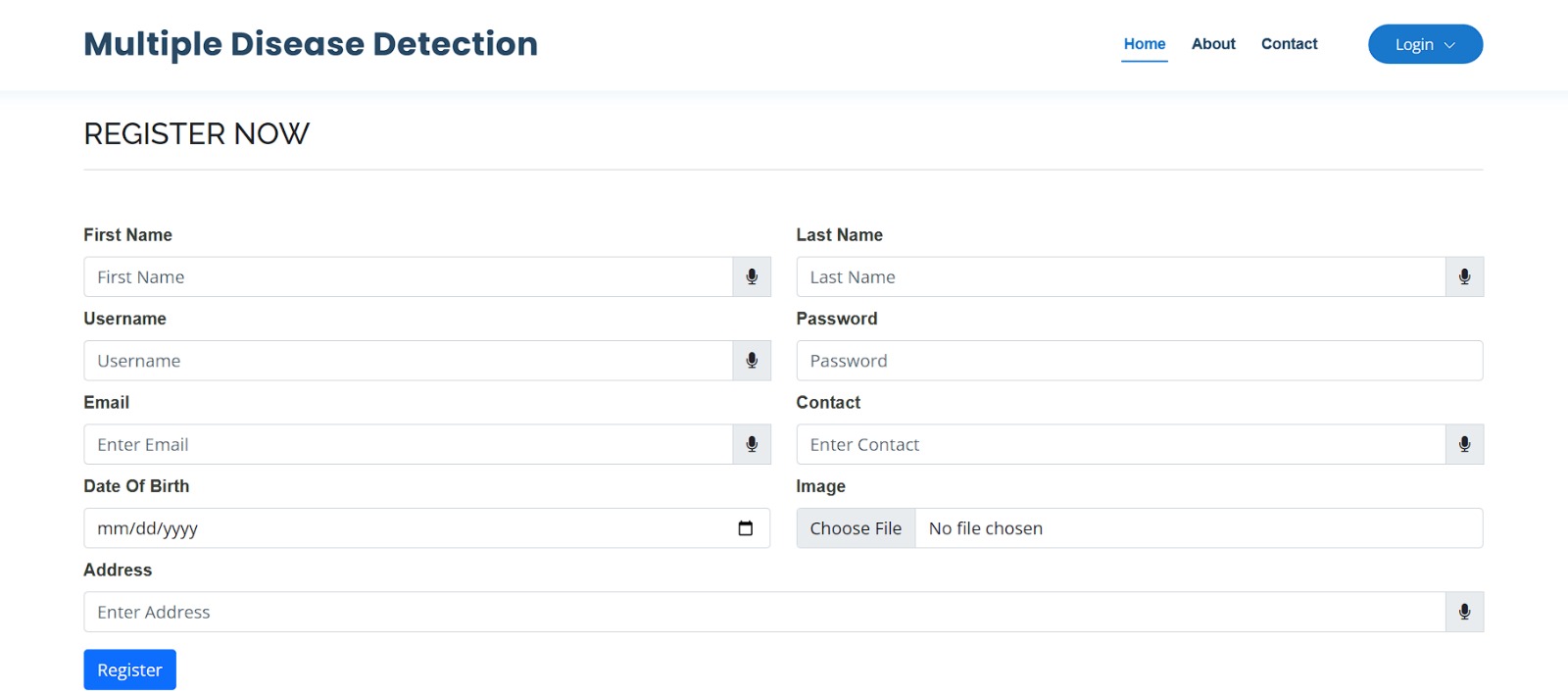
**5.3 Contact Page:** This is page that user can contact with authority or deverlopers if they got any bug/error is occurred and they can contact through the location also.

**Screenshot 5.3 Contact Page**

**5.4 Feedback Page:** User can give feedback from this page on the basis of interface, prediction, accuracy,etc

**Screenshot 5.4: Feedback Page**

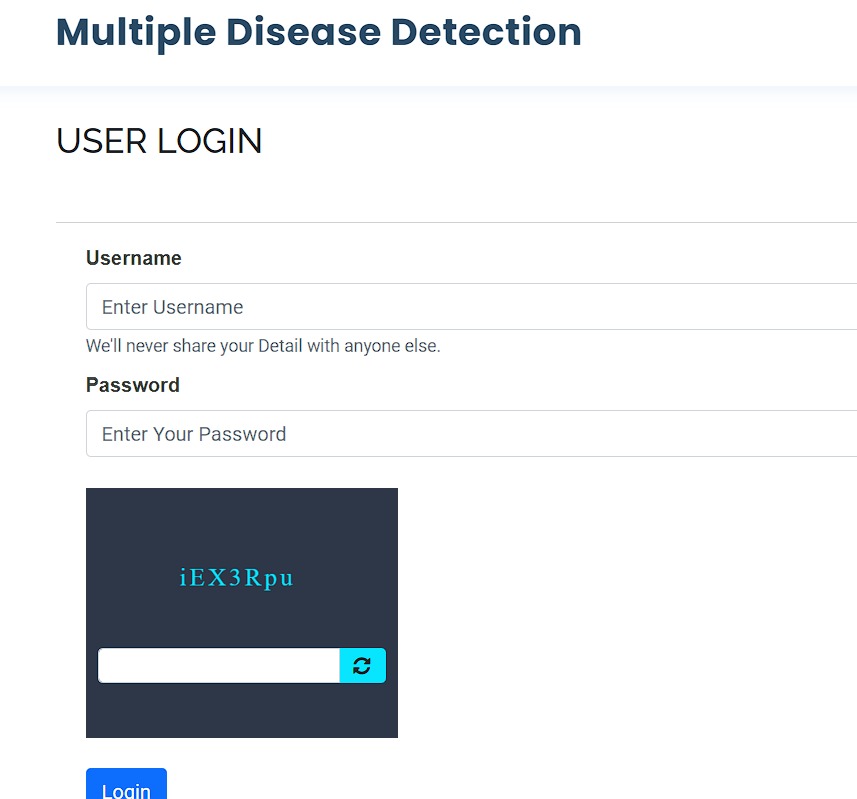
**5.5 Registration Page:** User first should be register himself from this page for the access of that application or becoming part of administration.



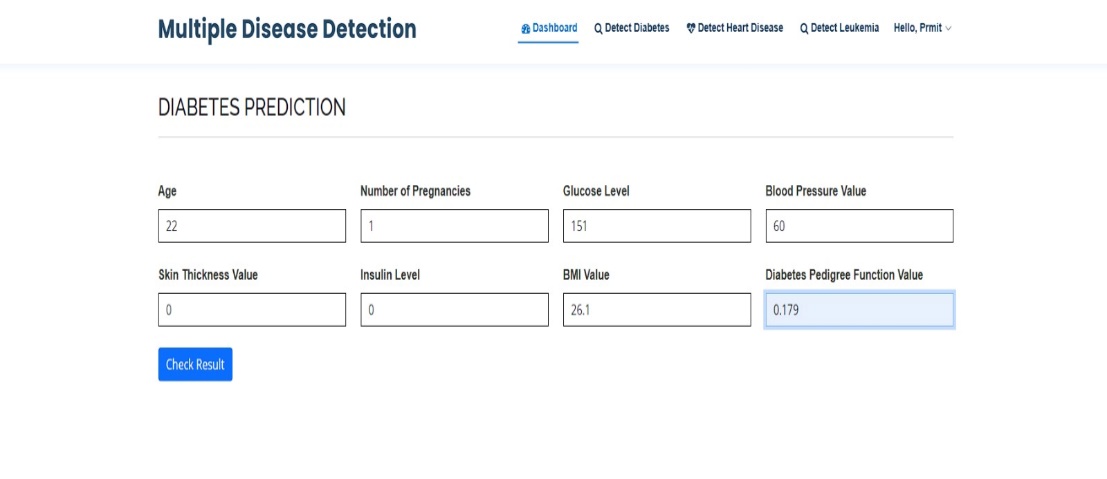
**Screensho**t **5.5: User Registration Page**

**5.6 Login Page:** From this page administrative person can log in to the system and can monitor the system.

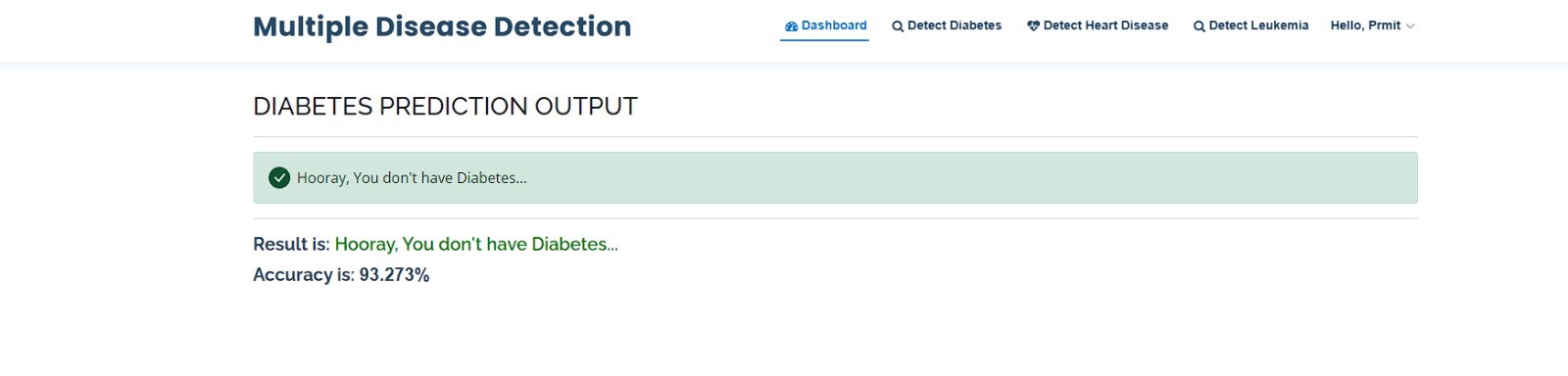
**s**

****

**Screenshot 5.6: User Login Page**

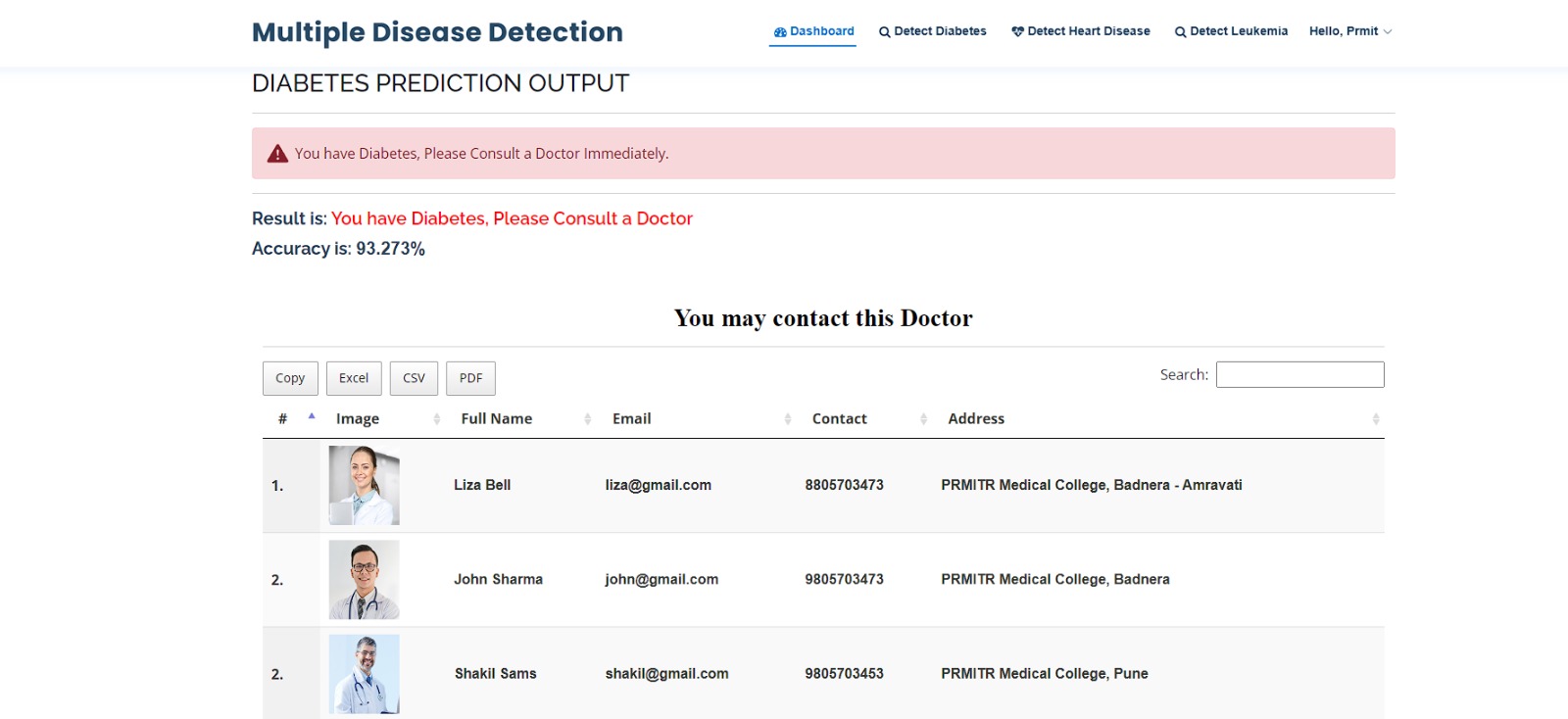
**5.7 Diabetes Prediction Page:** Here user can check his Diabetes ouput on the basis of data.

**Screenshot 5.7: Diabetes Prediction Page**

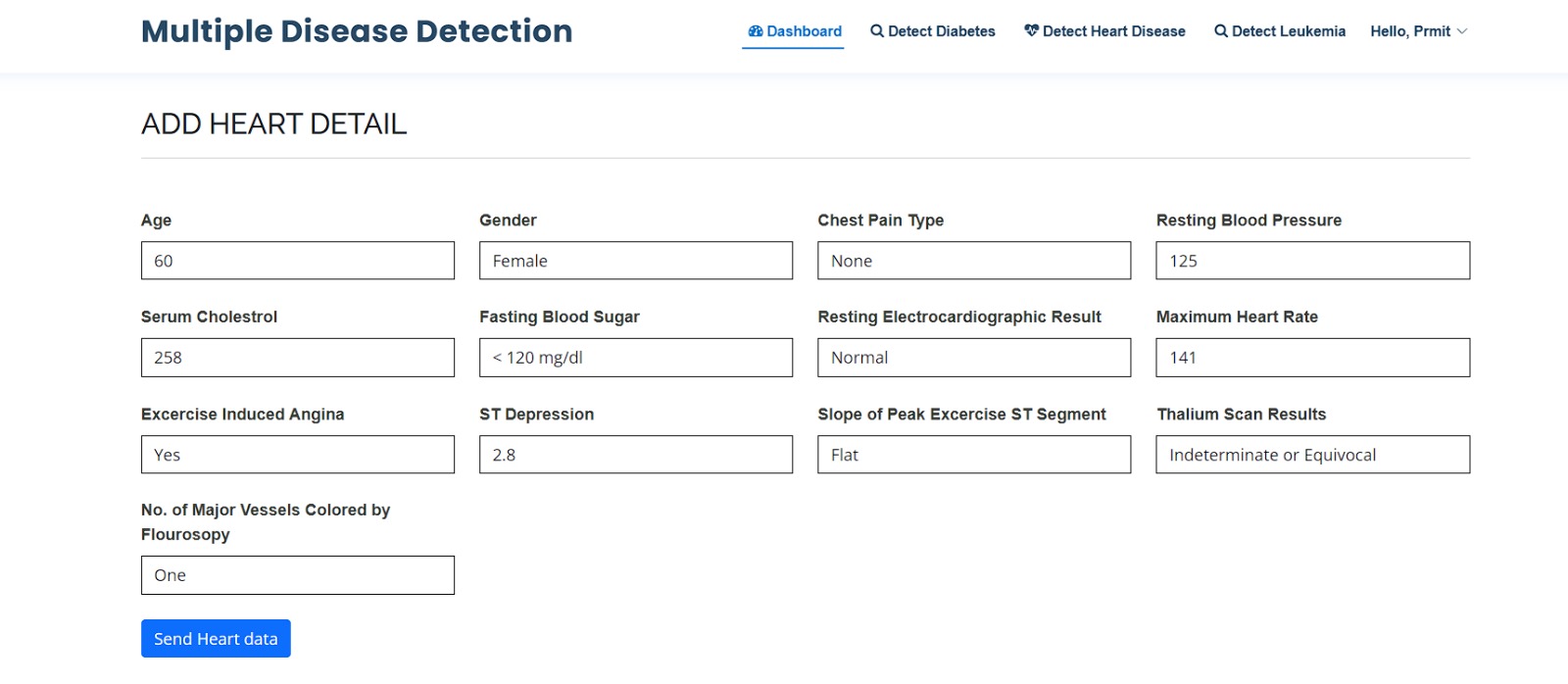
**5.8 Diabetes Output Page For Healthy Person:** After adding input data user got result, user is healthy person or not as shown in screenshot.

**Screenshot 5.8: Diabetes Output Page For Healthy Person**

**5.9 Diabetes Output Page For Unhealthy Person:** If user got positive result in diabetes then it is show in red letters and below that we are suggesting doctors list with contact details and address details, so user can contact them.

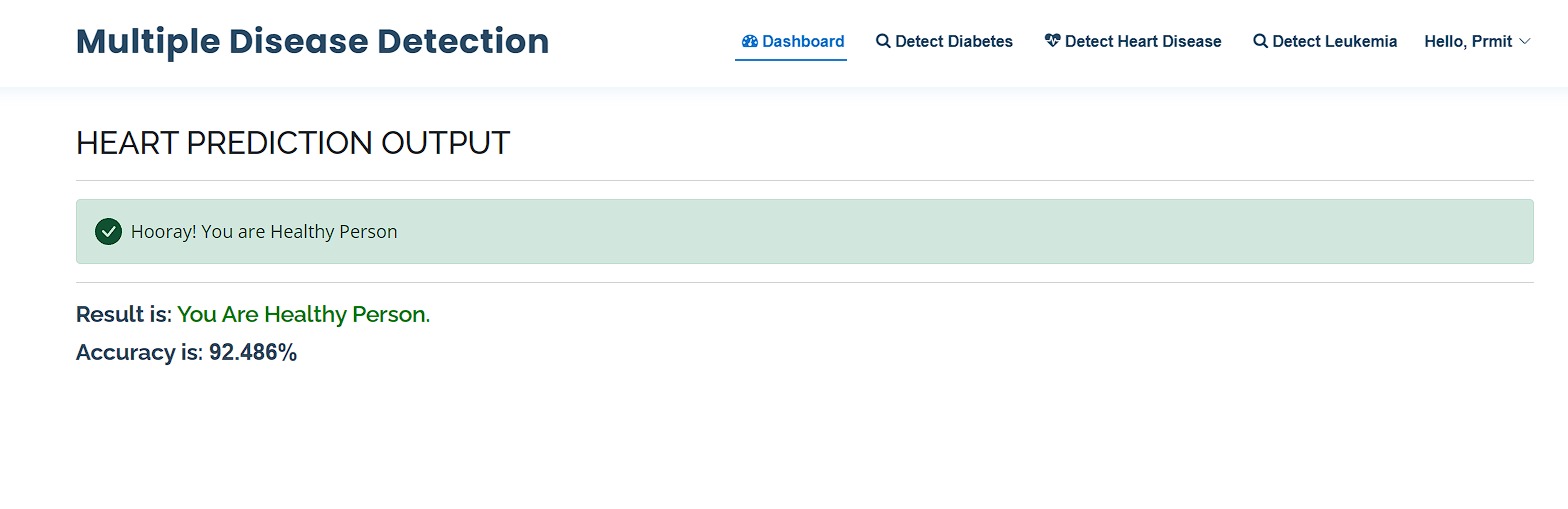


**Screenshot 5.9: Diabetes Output Page For Unhealthy Person**

**5.10 Heart Disease Prediction Page:** Here user can check his Heart Disease ouput on the basis of data.

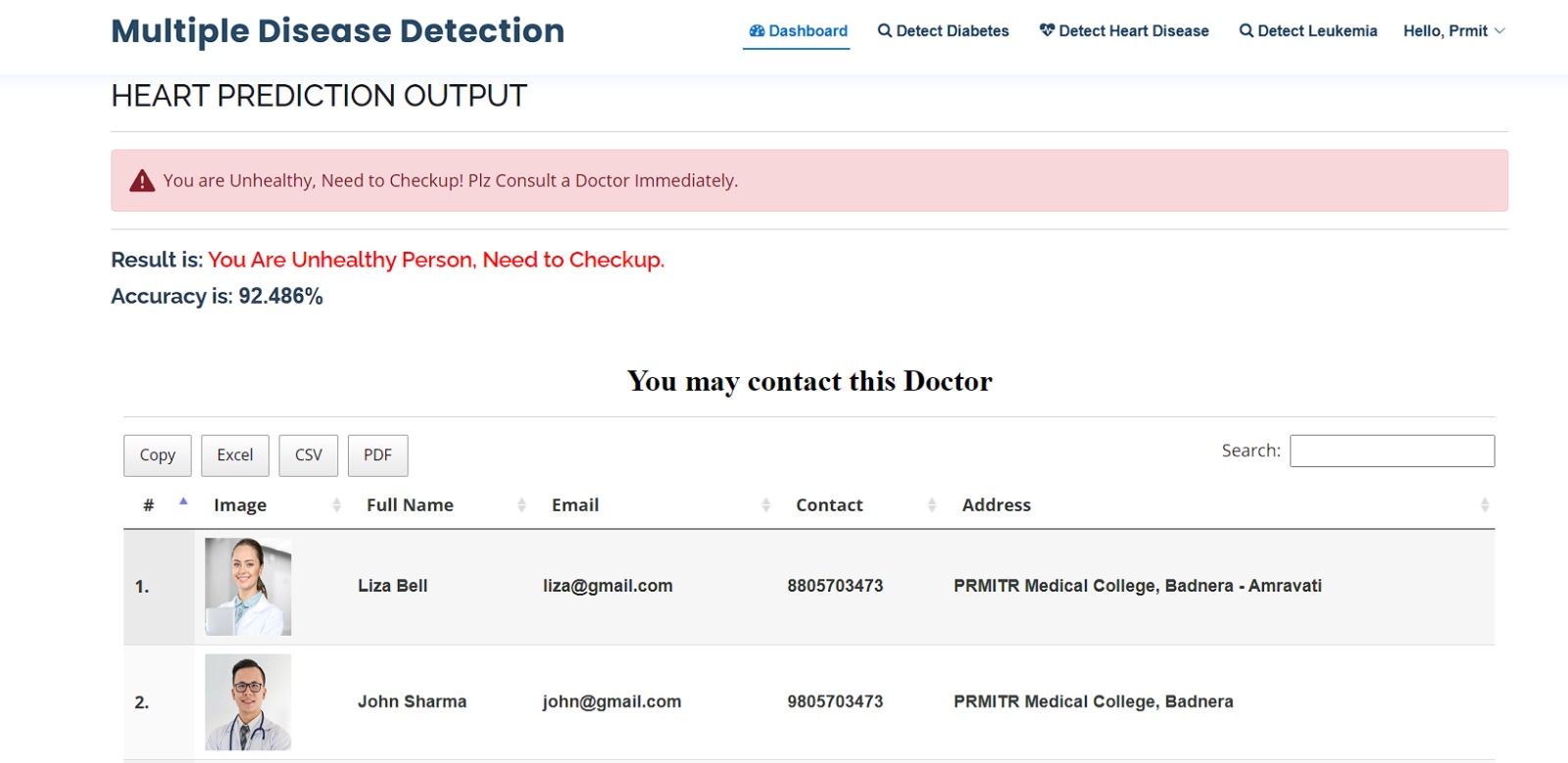
**Screenshot 5.10: Heart Disease Prediction Page**

**5.11 Heart Disease Output Page For Healthy Person:** After adding input data user got result, user is healthy person or not as shown in screenshot.



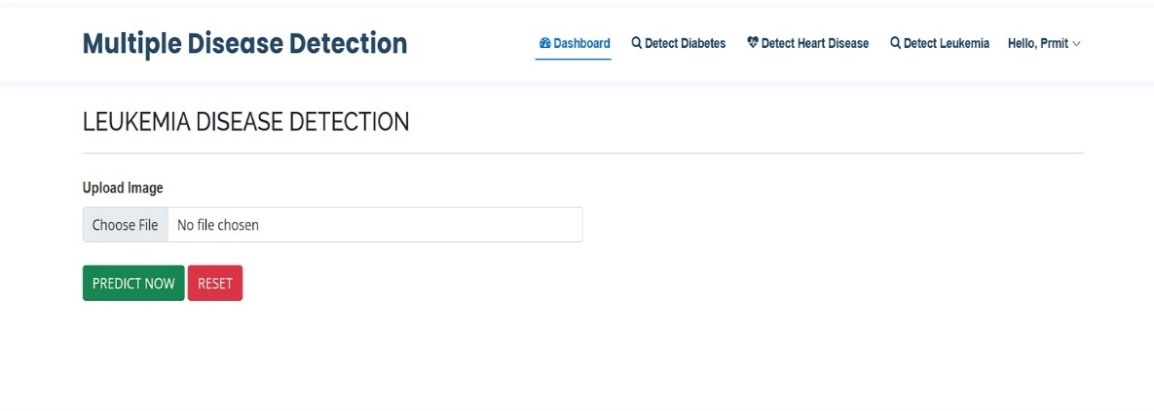
**Screenshot 5.11: Heart Disease Output Page For Healthy Person**

**5.12 Heart Image Output Page For Unhealthy Person:** If user got positive result in Heart disease then it is show in red letters and below that we are suggesting doctors list with contact details and address details same as diabetes, so user can contact them.

****

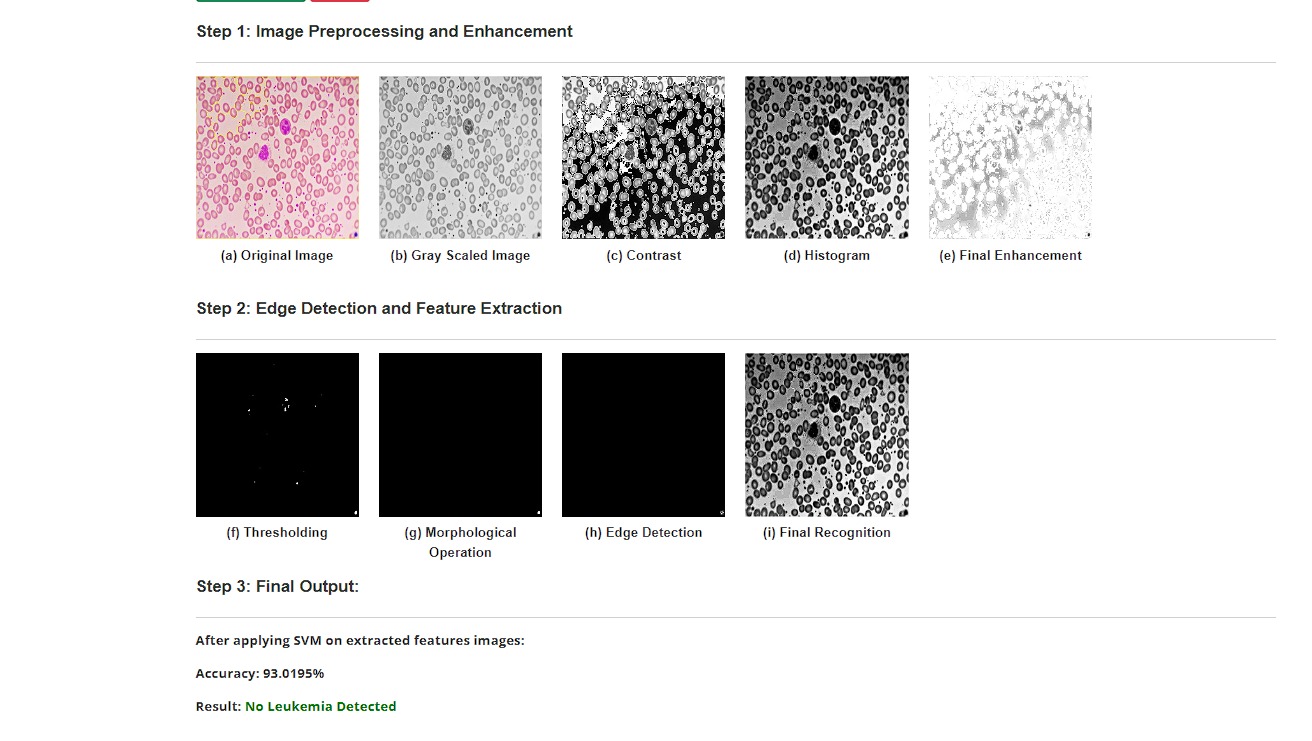
**Screenshot 5.12: Heart Disease Output Page For Unhealthy Person**

**5.13 Leukemia Prediction Page:** Here user can check his Leukemia ouput on the basis of data. In that we are providing input in image format.



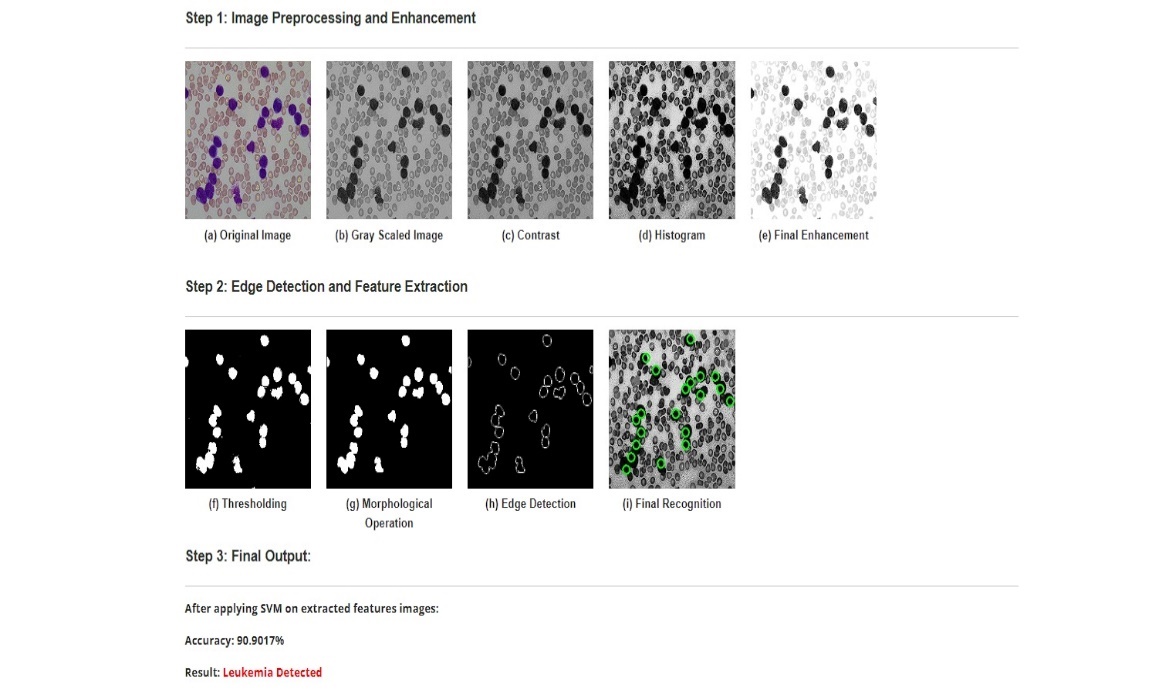
**Screenshot 5.13: Leukemia Disease Detection Page**

**5.14 Leukemia Output Page For Healthy Person:** After adding input data image, user got result Leukemia is detected or not as shown in screenshot.



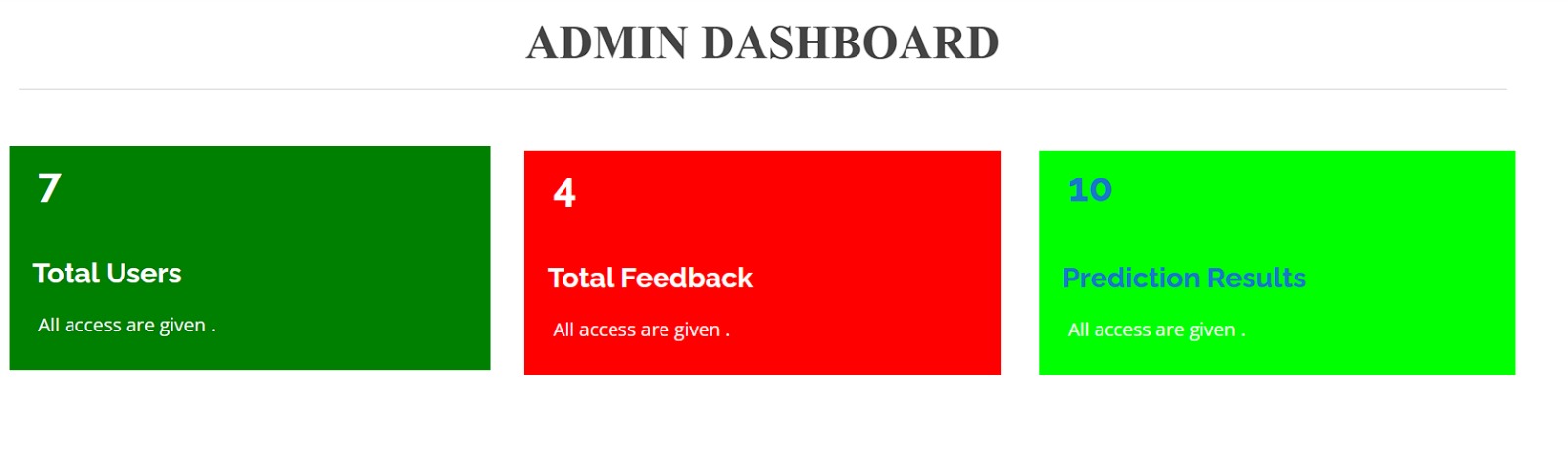
**Screenshot 5.14: Leukemia Detection Page Output For Healthy Person**

**5.15 Leukemia Output Page For Unhealthy Person:** If user got positive result in Leukemia then it is show in red letters and below that we are suggesting doctors list with contact details and address details same as above diseases so user can contact them.

****

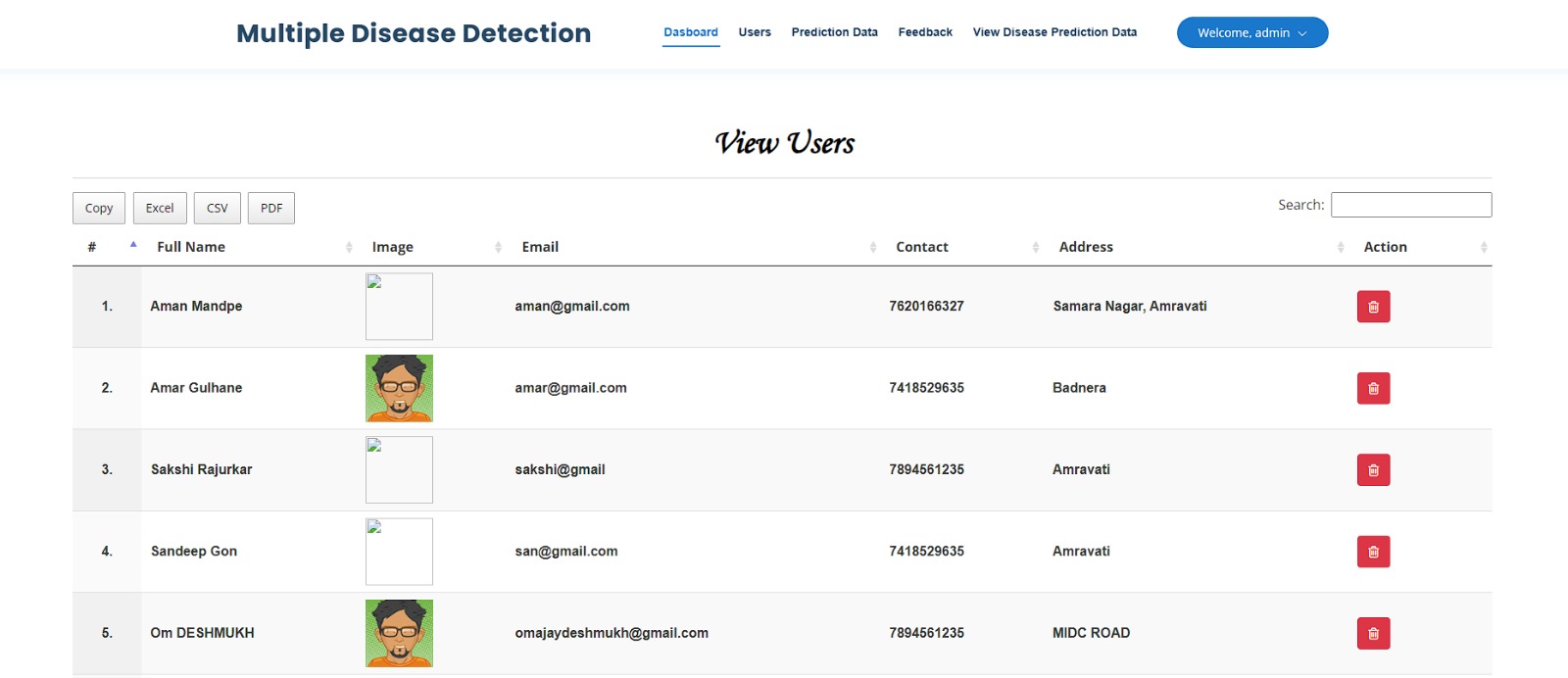
**Screenshot 5.15: Leukemia Detection Page Output For Unhealthy Person**

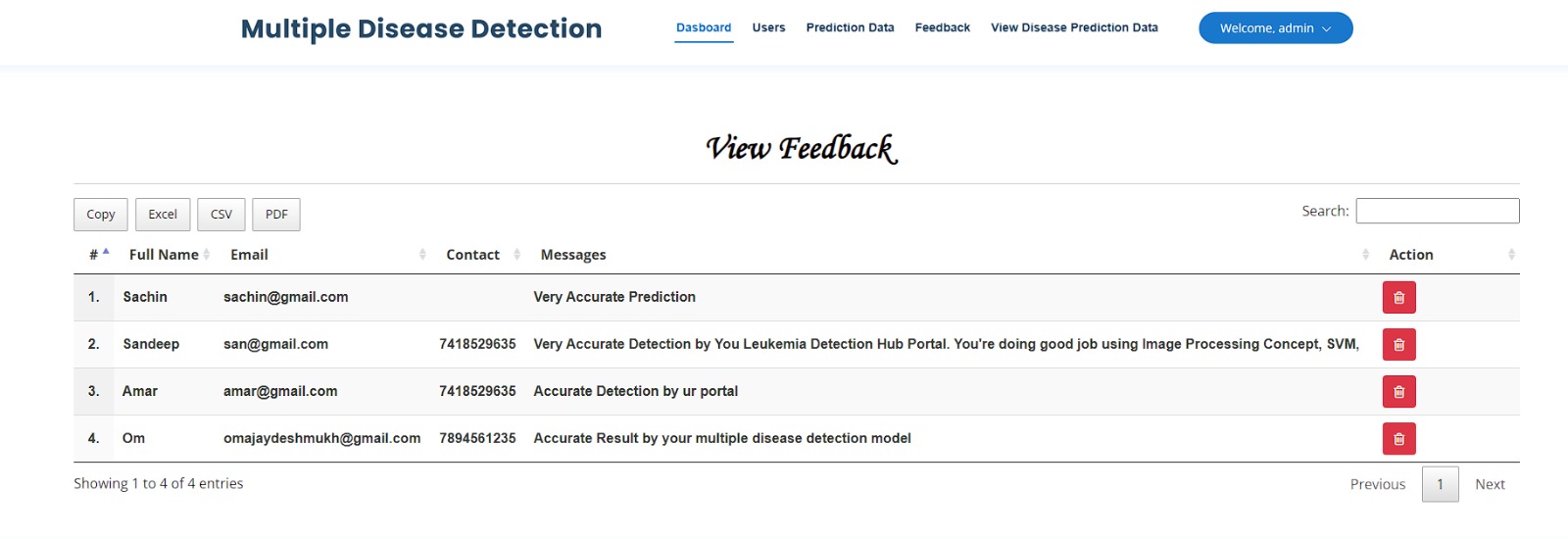
**5.16 Admin Page:** It is a Admin panel where only authority have access and they only check that three categories Total users, Total Feedback and Prediction Results as show in below screenshot.



**Screenshot 5.16: Admin Dashboard Page**

**5.17 Vier User Page:** The Admin Authirity can check total numbers of users in system in admin dashboard.

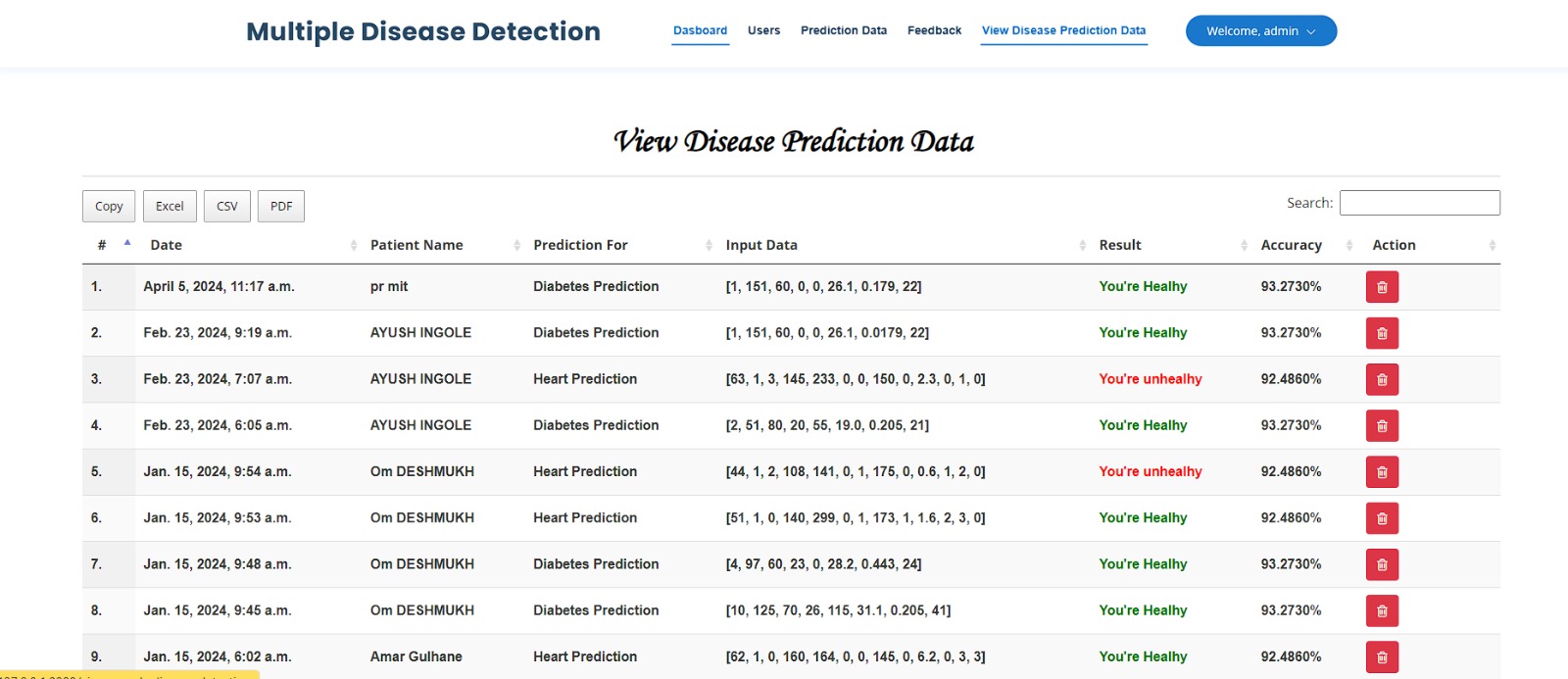


**Screenshot 5.17: View User Page**

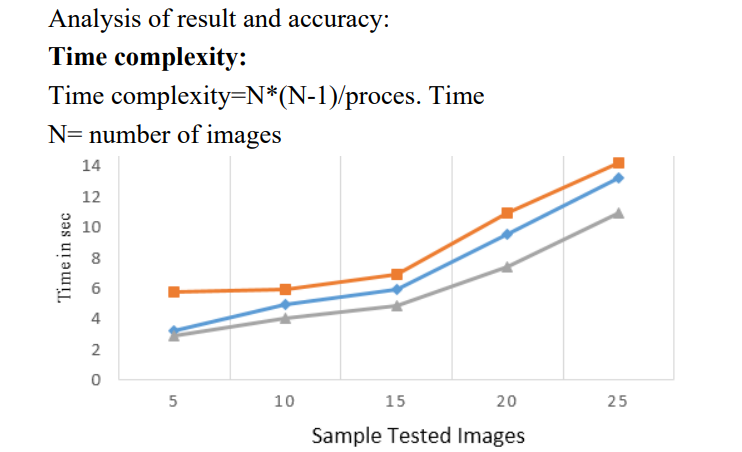
**5.18 View Feedback Page:** The Admin Authirity can check total no. of feedback of users in system in admin dashboard.

**Screenshot 5.18: View Feedback Page**

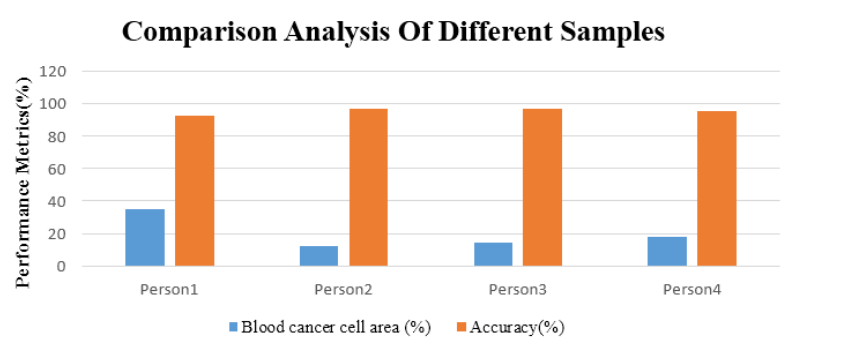
**5.19 All Prediction Data Page:** The Admin Authirity can check all no. of predicted result of users in system in admin dashboard.

****

**Screenshot 5.19: All Prediction Data Page**



**Figure 5.1:** Result & Accuracy



**Figure 5.2:** Analysis of Different Sample

**CHAPTER 6**

**CONCLUSION**

In this study, we have introduced a novel approach for the Multiple Disease Prediction By Using Machine Learning through the integration of image processing techniques, Feature Extraction, Logistic regression and a Support Vector Machine (SVM) model. Our model exhibits a commendable accuracy of 92% in detecting immature leukocytes, aligning itself with the current state-of-the-art methodologies.

The utilization of machine learning for the prediction of multiple diseases, including diabetes, heart disease, and leukemia, represents a pivotal advancement in the realm of healthcare. Through the amalgamation of diverse datasets encompassing clinical records, genetic information, imaging data, and lifestyle variables, predictive models have demonstrated remarkable efficacy in identifying individuals at risk of developing these debilitating conditions.

The future outlook for disease prediction using machine learning is highly promising, offering a plethora of opportunities across various domains. Advanced algorithms, such as deep learning architectures, enable the extraction of intricate patterns and correlations from vast datasets, thereby enhancing the accuracy and reliability of predictive models. Furthermore, the integration of real-time monitoring technologies and wearable devices facilitates proactive health management, enabling early detection of disease symptoms and timely intervention.

In the results section, we analyzed the dataset, detailing its characteristics and the preprocessing steps taken. We discussed feature selection methods and the rationale behind selecting machine learning algorithms. Results included model performance metrics like accuracy, precision, recall, and AUC-ROC. We compared different models, discussed feature importance, and addressed limitations. Suggestions for future research were also provided, concluding with the implications for healthcare practice and research.

**CHAPTER 7**

**ADVANTAGES, DISADVANTAGES AND FUTURE SCOPE**

**7.1 Advantages:**

1. **Early Detection:** Machine learning-based prediction models can identify the risk of multiple diseases at an early stage, allowing for timely intervention and treatment.
2. **Personalized Medicine:** By analyzing individual patient data, these models can tailor treatment plans to each patient's specific health profile, improving the effectiveness of medical interventions.
3. **Resource Optimization:** Predictive models can help healthcare providers allocate resources more efficiently by focusing on high-risk individuals or populations, thus reducing costs and improving healthcare delivery.
4. **Preventive Healthcare:** By identifying individuals at risk of developing certain diseases, machine learning models can facilitate preventive measures such as lifestyle modifications or screening programs, leading to better health outcomes.
5. **Improved Patient Outcomes:** Early detection and personalized treatment can lead to improved patient outcomes, including reduced morbidity, mortality, and complications associated with chronic diseases.

**7.2 Disadvantages:**

1. **Data Quality Issues:** Machine learning models heavily rely on the quality and representativeness of the input data. Inaccurate or incomplete data can lead to biased predictions and compromised model performance.
2. **Data Privacy Concerns:** Healthcare data often contain sensitive information about individuals' health conditions and medical history. Maintaining patient privacy and adhering to data protection regulations (such as HIPAA in the United States) pose significant challenges in data collection, sharing, and analysis.
3. **Algorithm Bias and Fairness:** Machine learning algorithms may exhibit bias towards certain demographic groups or population subgroups, leading to disparities in disease prediction and healthcare outcomes. Ensuring algorithm fairness and mitigating bias require careful attention to model training data and evaluation metrics.
4. **Model Interpretability:** Some machine learning algorithms, such as deep learning neural networks, are inherently complex and lack interpretability. Understanding the underlying factors driving disease predictions may be challenging, limiting the trust and acceptance of these models by clinicians and patients.
5. **Generalization Limitations:** Machine learning models trained on specific datasets may struggle to generalize to new or unseen data, especially if the distribution of the data changes over time or across different populations. Ensuring the robustness and generalizability of predictive models is crucial for their real-world applicability.

**7.3 Future Scope:**

Predicting multiple diseases using machine learning holds significant potential for revolutionizing healthcare by enabling early detection, personalized treatment plans, and improved patient outcomes. Integrating various types of medical data including electronic health records (EHRs), genomic data, medical imaging, wearable device data, and lifestyle factors will be crucial. Standardizing these diverse data sources will ensure compatibility and facilitate accurate disease prediction models. Combining different types of data (e.g., clinical, genetic, environmental) through advanced data fusion techniques will enhance the predictive power of models. For example, combining genetic information with clinical data can provide insights into disease predisposition and prognosis. Deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can effectively extract complex patterns from large-scale medical data.

Feature selection methods can help identify the most informative variables, improving model interpretability and efficiency. Incorporating disease prediction models into CDSS can assist healthcare providers in making evidence-based decisions at the point of care. These systems can provide clinicians with actionable insights, risk assessments, and personalized recommendations for diagnosis and treatment.

The future scope of multiple disease prediction using machine learning is promising, with potential applications across various domains of healthcare. By leveraging advanced techniques and integrating diverse data sources, predictive models can transform disease management, improve patient outcomes, and drive progress towards precision medicine.

**REFERENCES**

1. Sai, P. P., & Reddy, C. , “Heart disease prediction using ANN algorithm in data mining,” International Journal of Computer Science and Mobile Computing, 2022.
2. S. Ghane, N. Bhorade, N. Chitre, B. Poyekar, R. Mote and P. Topale, "Diabetes Prediction using Feature Extraction and Machine Learning Models," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2021.
3. Gulpreet Kaur Chadha, Aakarsh Srivastava, Abhilasha Singh, Ritu Gupta, Deepanshi Singla, “An Automated Method for Counting Red Blood Cells using Image Processing,” ScienceDirect , 2022, 769–77.
4. Science, C., & Faculty, G. M. , “Heart Disease Prediction Using Machine learning and Data Mining Technique,” Department of Computer Science and Engineering, Nirma University, Gujarat, India, 2023.
5. M. K. Hasan, M. A. Alam, D. Das, E. Hossain and M. Hasan, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers," in IEEE Access, vol. 8,2022.
6. George, J., Sobha, T., Ashok, A., & Eldhose, J. , “An efficient image processing technique to count red blood cells,” International Journal of Engineering and Research, 2020.
7. A, A. S., & Naik, C. , “Different Data Mining Approaches for Predicting Heart Disease,” International Journal of Innovative in Science, 2021.
8. Beyene, C., & Kamat, P. , Survey on “Prediction and Analysis the occurrence of heart disease using data mining techniques,” International Journal of Pure and Applied Mathematics, 2021.
9. Irwan Rahadi, “Red blood cells and white blood cells detection by image processing,” [Journal of Physics: Conference Series](https://iopscience.iop.org/journal/1742-6596), [Volume 1539](https://iopscience.iop.org/volume/1742-6596/1539), [The 5th Hamzanwadi International Conference of Technology and Education, Lombok, Indonesia](https://iopscience.iop.org/issue/1742-6596/1539/1), 2020.
10. P. Sonar and K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches," 2020 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021.
11. Sultana, M., Haider, A., & Uddin, M. S. , “Analysis of data mining techniques for heart isease prediction,” 3rd International Conference n Electrical Engineering and Information and Communication Technology, 2022.
12. Soni, J., Ansari, U., & Sharma, D., “Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers,” Heart Disease, 3(6), 2385–2392, 2021.
13. Reddy Bhavana , Prajjwal Srivastava , Sangeetha N , Rashmi Reddy “ Review on Identification of Red Blood Cells by Image Processing,” IEEE International Journal of Engineering Research & Technology (IJERT), 2021.
14. D. Shetty, K. Rit, S. Shaikh and N. Patil, "Diabetes disease prediction using data mining," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2023
15. Kirmani, M, “Cardiovascular Disease Prediction using Data Mining Techniques,” Oriental Journal of Computer Science and Technology, 2021, 10(2), 520–528.

**FINANCIAL SHEET**

As per our analysis, the project we have been working on does not require any hardware and can be implemented with only software. Since there was no requirement for hardware, we were able to keep the costs relatively low and utilize our resources effectively.

|  |  |
| --- | --- |
| Total hours spent | 40 hours |
| Rate per Hour | Rs. 341 |
| Total cost | Rs 13,640 |

Our group has worked tirelessly to ensure that the software is functioning optimally, and we are confident that it will be able to deliver the desired results. We believe that the software solution we have developed is both cost-effective and efficient, and we are excited to see the positive impact it will have on our organization.