STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION

A PROJECT REPORT

submitted by

ANUPAMA JEYASHREE S (412519205012)

PARVATHI PRIYA NANDANA K M (412519205098)

LAVANYA M (4125192050)

ANITHASHREE V (412519205011)

NOV 2022

TABLE OF CONTENTS

8. **TESTING**
    1. Test Cases
    2. User Acceptance Testing
9. **RESULTS**
    1. Performance Metrics
10. **ADVANTAGES & DISADVANTAGES**
11. **CONCLUSION**
12. **FUTURE SCOPE**
13. **APPENDIX**

    Source Code

    GitHub & Project Demo Link

LIST OF FIGURES

1. INTRODUCTION

1.1 Project Overview

Statistical Machine Learning Approaches to liver disease prediction is used to find the liver disease in a human being. Cirrhosis is a leading cause of mortality and morbidity across the world. It is the 11th leading cause of death and 15th leading cause of morbidity, accounting for 2.2% of deaths and 1.5% of disability. Among the various liver disease types, Chronic Liver Disease (CLD) caused 1.32 million deaths in 2017, approximately two-thirds among men and one-third among women.

1.2 Purpose

The purpose of this project is to identify whether a user has liver disease or not based upon the data entered as input. Based on the inputs entered by the user, a result is displayed**.** If the user has been diagnosed with any form of liver disease, then it is highly recommended that they seek medical attention immediately. Our project's objective is to analyze data from liver patients with a focus on the correlations between a vital list of liver enzymes, proteins, age and gender using them to try and predict the likeliness of liver disease. We are creating a model in this instance by utilizing different machine learning algorithms to identify the most accurate model. and incorporate into web applications built with flask. By entering certain parameters into the web application, users can predict the disease.

2. LITERATURE SURVEY

2.1 Existing Problem

[1] In this system they described how to predict risk of liver disease for a person, based on the blood test report results of the user using various machine learning algorithms. The final output was predicted based on the most accurate machine learning algorithm. Based on the accurate model they designed a system which asks a person to enter the details of his/her blood test report. Then the system uses the most accurate model which is trained to predict, whether a person has risk of liver disease or not.

[2] In the 21st-century, the issue of liver disease has been increasing all over the world. The overall percentage of death by liver disease is 3.5% worldwide. Chronic Liver disease is also considered to be one of the deadly diseases, so early detection and treatment can recover the disease easily. This research work is based on liver disease prediction using machine learning algorithms. Liver disease prediction has various levels of steps involved, pre-processing, feature extraction, and classification. In this s research work, a hybrid classification method is proposed for liver disease prediction. And Datasets are collected from the Kaggle database of Indian liver patient records. The proposed model achieved an accuracy of 77.58%.

[3] Machine learning (ML) utilizes artificial intelligence to generate predictive models efficiently and more effectively than conventional methods through detection of hidden patterns within large data sets. In this review, we examine the literature pertaining to machine learning in hepatology and liver transplant medicine. We provide an overview of the strengths and limitations of ML tools and their potential applications to both clinical and molecular data in hepatology. ML has been applied to various types of data in liver disease research, including clinical, demographic, molecular, radiological, and pathological data. We anticipate that use of ML tools to generate predictive algorithms will change the face of clinical practice in hepatology and transplantation.

[4] Data Mining technologies have been widely used in the process ofmedical diagnosis and prognosis, extensively. In this project, the patient data sets are analyzed for the predictability of the subject to have a liver disease based purely on a widely analyzed classification model. This System predict the same conclusive result with a higher rate of accuracy. J48 algorithm is considered to be a better performing algorithm when it comes to feature selection with an accuracy rate of 95.04%.

[5] Various kinds of pressure and unbalanced eating behaviors, along with alcohol inhalation and on-going toxic gases, etc, cause lever disease in patients.For this purpose, the type of data

mining algorithms can help medical doctors to diagnose patients in hospital. This paper analyzes meta learning algorithms to classify the Indian liver patient dataset. Adaboost, logitboost, Bagging and Grading meta learning algorithms are applied to this data set. Key role is played by Grading algorithm in shaping enhanced classification accuracy (Correct Classification Rate) of a data set.

2.2 Reference

[1] Rakshith D B, Mrigank Srivastava, Ashwani Kumar, Gururaj S P, “Liver Disease Prediction System Using Machine Learning Techniques”, IJERT, Vol-10, Issue 6, June 2021.

[2] Shaheamlung, Golmei & Kaur, Harshpreet. (2021). The Diagnosis of Chronic Liver Disease using Machine Learning Techniques. INFORMATION TECHNOLOGY IN INDUSTRY. 9. 10.17762/itii.v9i2.382.

[3] Spann A, Yasodhara A, Kang J, Watt K, Wang B, Goldenberg A, Bhat M. Applying Machine Learning in Liver Disease and Transplantation: A Comprehensive Review. Hepatology (Baltimore, Md.). 2020 Mar;71(3):1093-105.

[4] Durai V, Ramesh S, Kalthireddy D. Liver disease prediction using machine learning. Int. J. Adv. Res. Ideas Innov. Technol. 2019;5(2):1584-8.

[5] Pasha M, Fatima M. Comparative Analysis of Meta Learning Algorithms for Liver Disease Detection. J. Softw.. 2017 Dec 1;12(12):923-33

2.3 Problem Statement Definition

Statistical Machine Learning Approaches To Liver Disease Prediction

Liver Disease prevent the liver's normal function. An effective task that can assist clinicians in quickly diagnosing the disease is the early prediction of liver disease using classification algorithms. This project's main goal is to compare the predicted accuracy of different categorization algorithms by analysing their parameters. A core list of liver enzymes, proteins, age, and gender are the focus of this project's analysis of patient data from the liver, which aims to forecast the likelihood of liver illness. ML algorithms are new techniques to handle many hidden problems in medical data sets. This approach can help healthcare management and professionals to explore better results in numerous clinical applications, such as medical image processing, language processing, and tumor or cancer cell detection, by finding appropriate features.

3. IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas

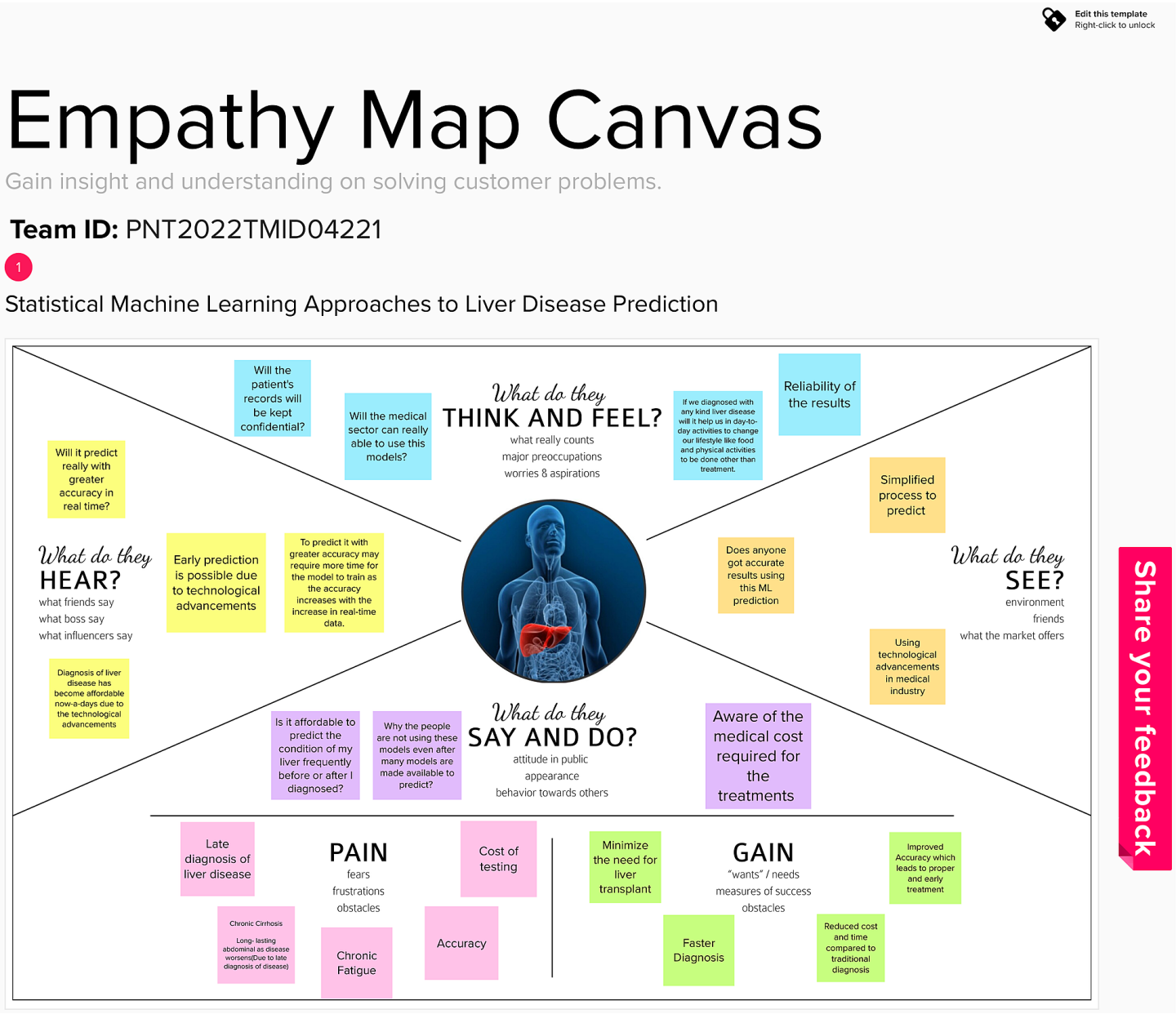


FIG 3.1 Empathy Map Canvas

3.2 Ideation and Brainstorming

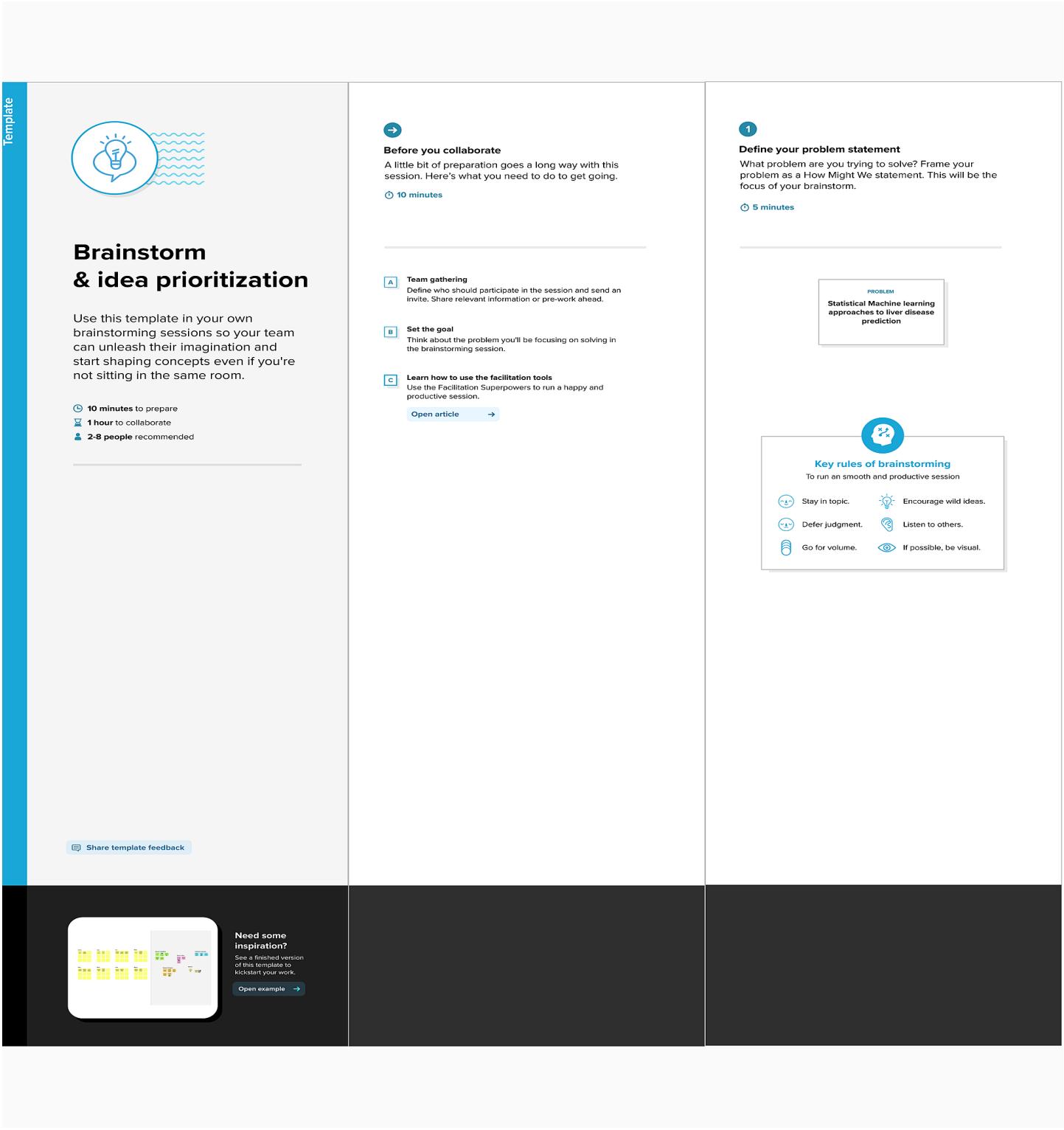


FIG 3.2.1 Brainstorming Ideas - Part one

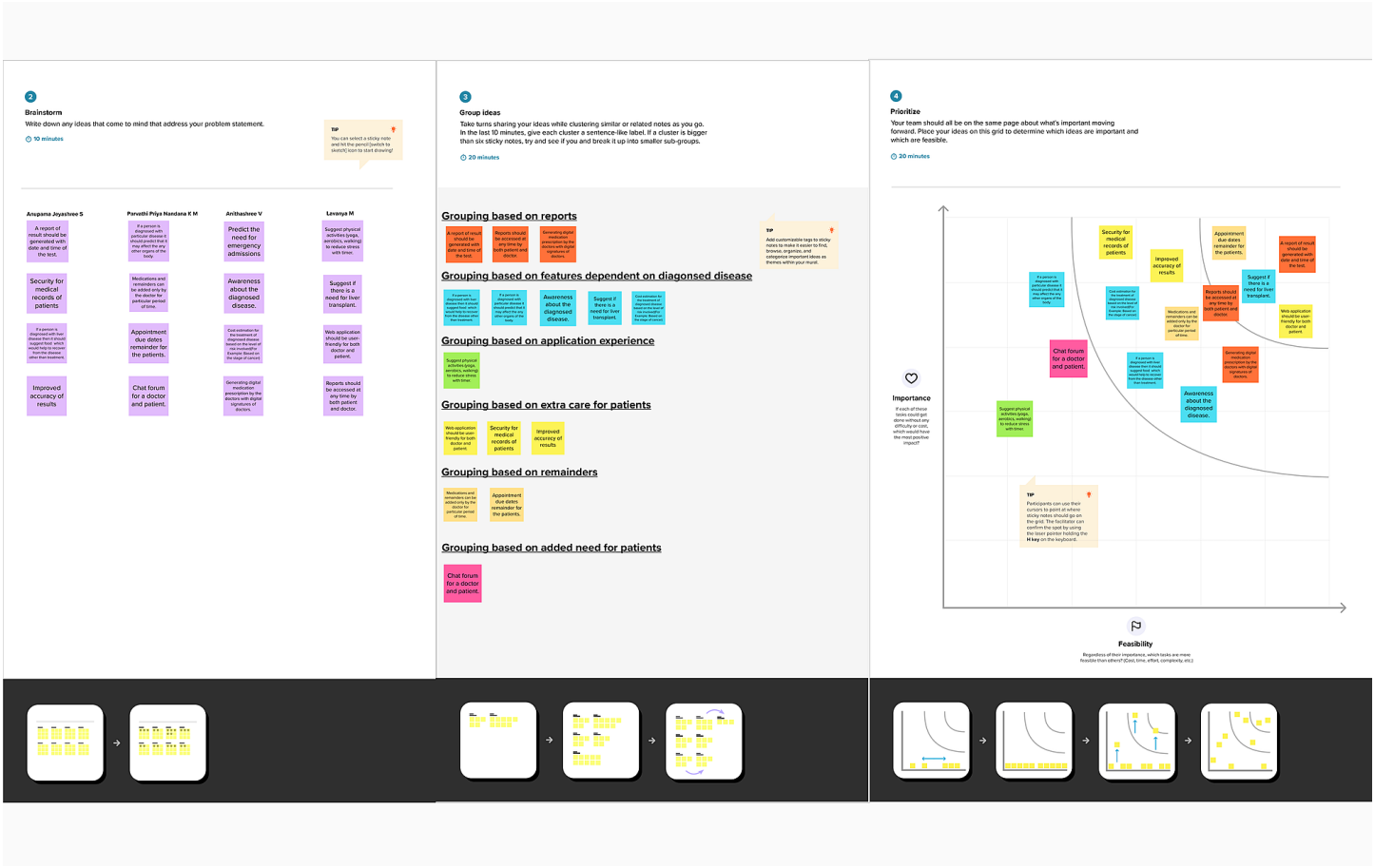


FIG 3.2.2 Brainstorming Ideas - Part two

3.3 Proposed Solution

S. No	Parameter	Description
1.	Problem statement	Around 2 million people die from liver diseases each year in the world, including 1 million from cirrhosis complications, 1 million from viral hepatitis, and 1 million from hepatocellular carcinoma. A liver performs many metabolic functions, including nutrient processing and distribution. Everyone in ten peopleget affected by liver diseases due to the adverse effects of drugs or autoimmune reactions. Commonliver infections are cirrhosis, liver cancer, Wilson disease, hepatitis A and B. Acute hepatitis continues to be largely caused by drug-induced liver injury, whileviral hepatitis remains highly prevalent.
2.	Idea description	Due to this, it is necessary to deal with this issue in advance. In order to avoid the loss of life due to liver diseases, early detection is crucial. Technological advances have led to many applications that predict liver diseases in different technical fields. However, there are no applications that provide the most accurateresults. Therefore, the goal of the project is to identify liver disease in patients by taking list of age, proteins, and enzymes into consideration. The main model will be developed using various statistical algorithms and will be attached with web application to provide accurate results to the patients. Usingcertain datasets, the model will be trained with algorithms, and it will be processed to predict the type of liver disease.

3.	Novelty / uniqueness	The objective of this project is to determine the best algorithm for determining liver disease at an earlystage by considering the values of different machine learning algorithms and analysing their results. Considering those machine learning algorithms like Naïve Bayes, K-Nearest Neighbour (KNN) etc., it is also necessary to consider the input values entered by the user.By analysing both the accuracy results for the model and the machine learning algorithms, we can determine which algorithm is most effective inpredicting liver disease.
4.	Social impact / customer satisfaction	Liver disease is one of the key causes of high numbers of deathsin the country and is considered a life- threatening disease, not just anywhere, but worldwide. Liver disease can also impact people earlyin their life. The application will have the option to predict liver infection beforeand advise the wellbeing condition. These methods can reduce many of the limitations that occur in healthcare associated with inaccuracy in diagnoses, missing data, cost, and time.The application of the ML methods can help reduce the total burden of liver disease on publichealth worldwide by improving the recognition of risk factors and diagnostic variables. More importantly, forchronic liver disease, detecting liver diseaseat earlier stages orin hidden cases by ML could decrease liver-related mortality, transplants, and/or hospitalizations. Early detection improves prognosis since treatment can be given before the progression of the disease to later stages. Invasive tests, such as biopsy, would occur lessin this case as well. Although this study focused on hepatitis and chronic liver disease variables for MLtraining, it canbe hypothesized that the methods can beused to distinguish other types of liver disease from healthy individuals.

3.4 Problem Solution Fit

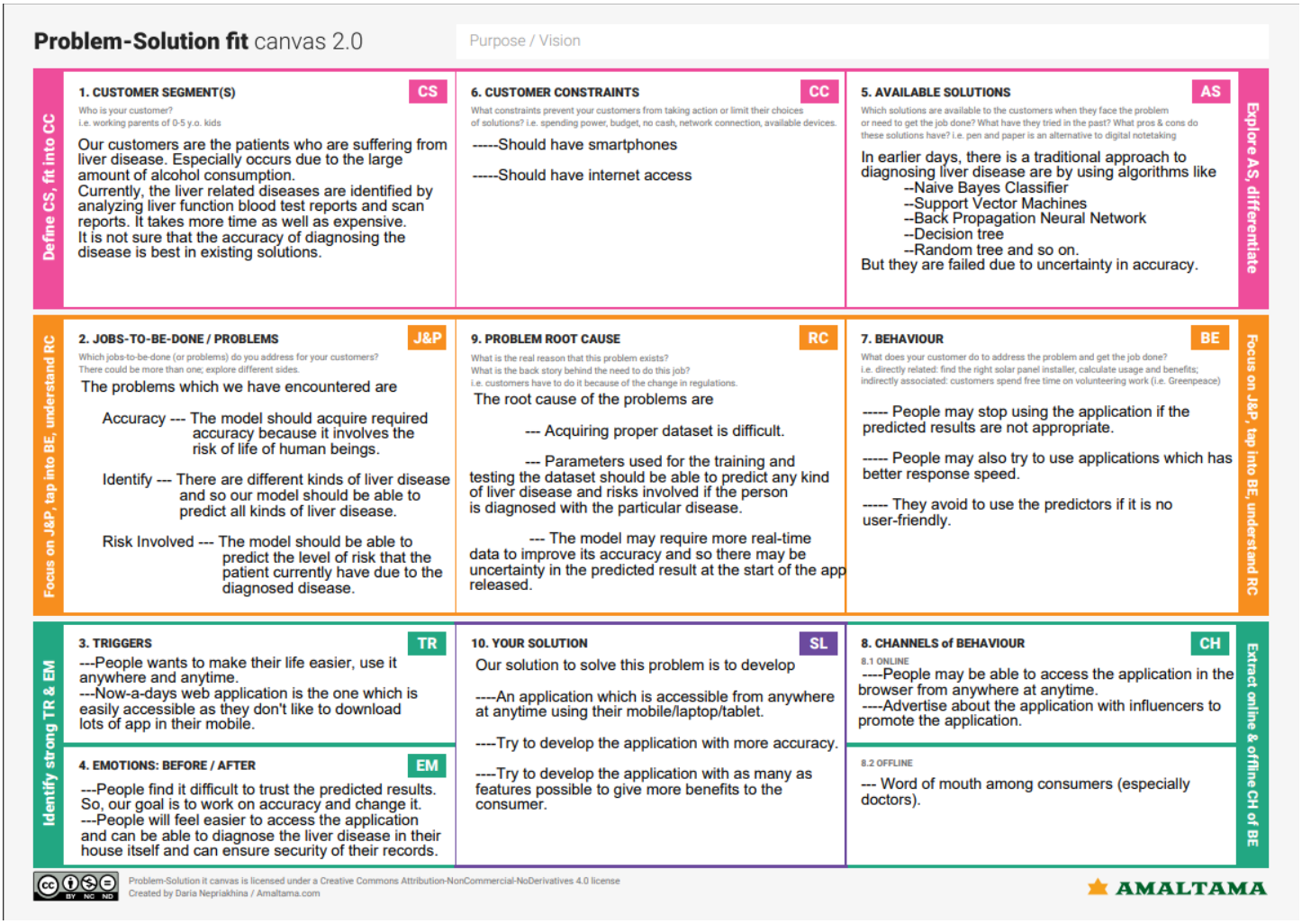


FIG 3.4 Problem Solution Fit

4. REQUIREMENT ANALYSIS

4.1 Functional Requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR -1	User Registration	The user registration process includes the creation of account through either email id or phone number with new password through the website.
FR - 2	User Login	The existing user can directly login to the website by giving the Login credentials.
FR - 3	Admin Login	The admin can login to the website where the admincan find the analysis of the predicted data.
FR - 4	Upload Image	The user can upload the scanned image of liver in the dropdown menu from various assets like (drop box, gallery etc.,)
FR-5	Enter Data	The user can enterthe required data from the scanned reports.
FR - 6	Prediction	The data entered by the user is tested with the trained model in the Watson studio to predict whether the user has liver disease or not.
FR-7	Display	The result will be displayed in theapplication.
FR-8	Report generation	The resultof the tested data will be generated as report in the form of PDF and stored in the user login and also it will automatically download to the user system.

4.2 Non-Functional Requirement

NFR No.	Non-Functional Requirement	Description
NFR - 1	**Usability**	The application can be easily accessible by any type of individuals, the aged individual and affected by liver disease can also use this toolfor Diagnosis.
NFR - 2	**Security**	Data security is important to store the customer data in the secured manner. The information should not be leaked outside.
NFR - 3	**Reliability**	Should provide results with more accuracy compared with existing solutions and consume less time than other existing solutions.
NFR - 4	**Performance**	The ability of MachineLearning is to perform pattern recognition by creating complex relationships based on input data and then comparing it with performance standards is a big step also to diagnosis in short time.
NFR - 5	**Availability**	Healthcare affordability, standards, and accessibility is made much more easier using this platform and the application will be available to allkinds of users.
NFR - 6	**Scalability**	The application must hold stable even whenmultiple users are using it at the same times.

5.PROJECT DESIGN5.1 Data Flow Diagrams

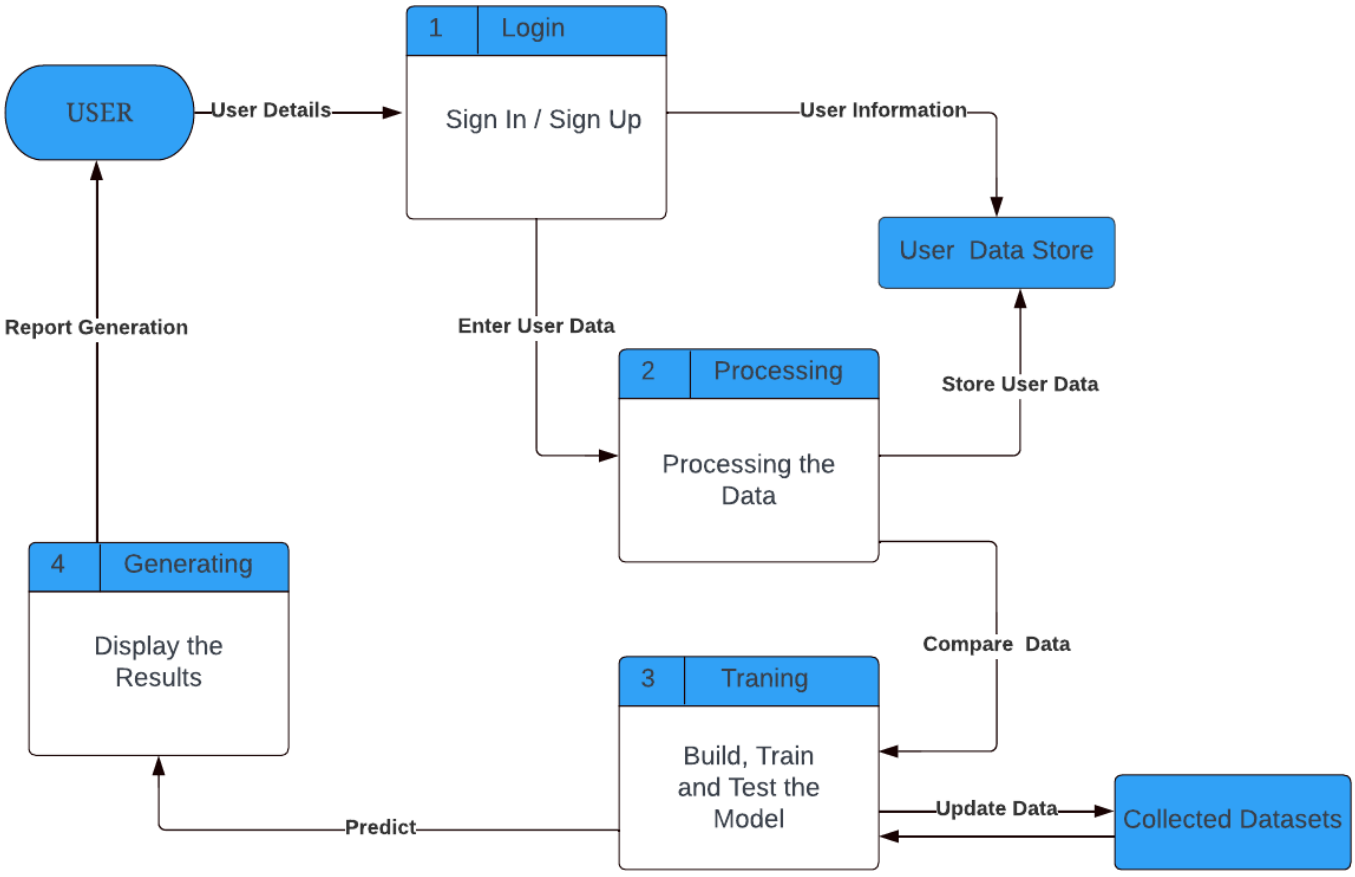


FIG 5.1 Data Flow Diagram

5.2 Solution and Technical Architecture

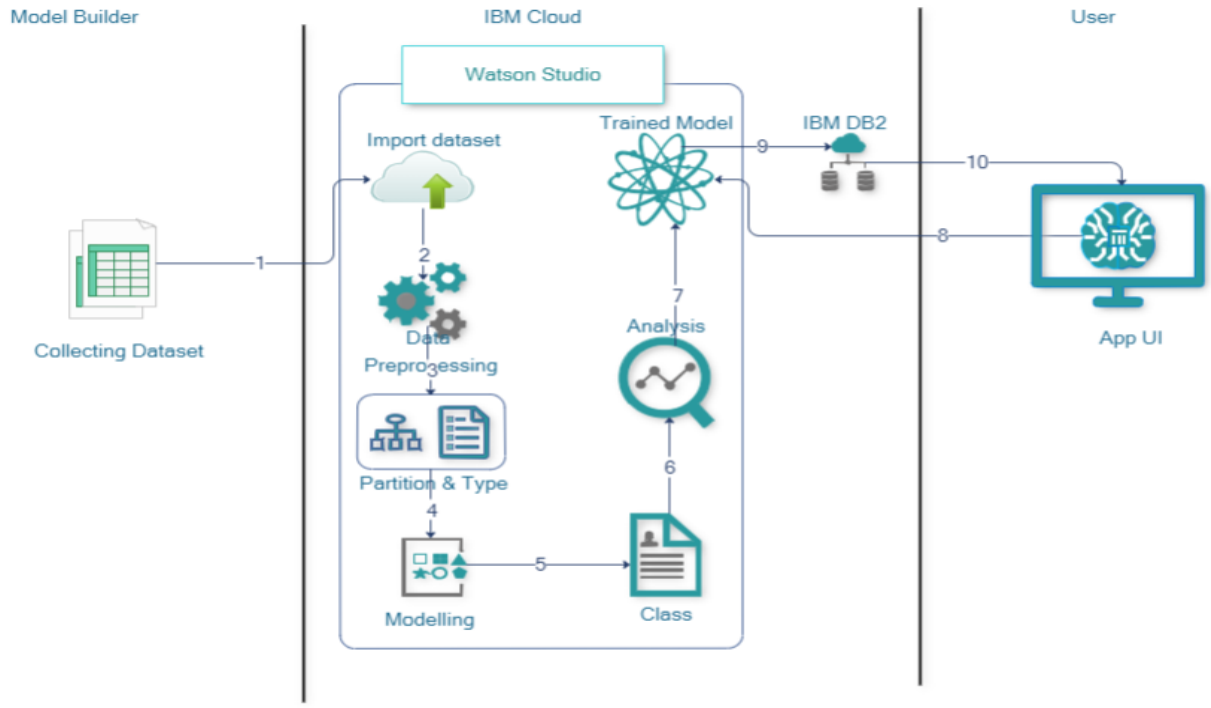


FIG 5.2 Technical Architecture

1. Import the collected dataset into the Watson studio.

2. Process the datasets as per the requirements.

3. Data pre-processing can be done by partitioning and Type Partitioning—splits data into separate subsets for training, testing and validation stages of model building Type—Specify field metadata and properties that are invaluable for modelling.

4. Build the model by splitting the sample based on the field that gives maximum information gain.

5. A class of data is derived from the dataset with maximum information gain.

6. Evaluate the ability of a model to generate accurate predictions.

7. Final model is developed based on evaluation.

8. User giving their data into the app.

9. Predict the results and store it into IBM DB2 cloud database.

10. Display the results to the user in the application.

Components & Technologies

S. No	Component	Description	Technology
1.	User Interface	User interacts with the system through the developed Web Application	HTML, CSS, Js, Flask
2.	Building Model	Pre-process the dataset, train the model using the train data and test the model with the test data and user input data as per performance metrics.	Python, Numpy, Scikit-learn, Tensorflow
3.	Fine tuning the model	Model is fine tuned to increase the accuracy of prediction	Optimizer, Tensorflow
4.	Navigation within Web UI	All the available features can be accessed from the dashboard.	Flask
5.	Cloud Database	Database Service on Cloud	IBM DB2
6.	File Storage	File storage requirements	IBM Block Storage
7.	External API	Login/Registration through Google Account	Google API
8.	Machine Learning Model	To detect Liver Disease using Machine Learning	SVM Algorithm, Xception, VGG19
9.	Cloud Infrastructure	Cloud Server Configuration	Cloud Foundry

Application Characteristics:

S. No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Flask micro web framework	Python, Numpy, Tensorflow, Scikit-learn, IBM Watson, Google API, Flask
2.	Security Implementations	With all aspects of the job including detecting malicious attacks, analysing the network endpoint protection and vulnerability assessment, Sign-in Encryption	IBM Cloud App ID Services
3.	Scalable Architecture	When we scale up the hardware capacity, the app can be able to handle the workload to scale up to the same degree.	IBM Cloud
4.	Availability	Available for all data size	IBM Cloud Service
5.	Performance	Can extend the storage according to our needs	Python, IBM Cloud

5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the Application.	I can receive Confirmation email & click confirm.	High	Sprint-1
	Login	USN-3	As a user, I can log into the application by entering email&password	I can access the dashboard	High	Sprint-1
	Dashboard	USN-4	As a user, I must enter my details	I can retrieve information anywhere	Medium	Sprint-2

	Dashboard	USN-5	As a user, I can navigate through different pages usingthe dashboard	I can view various pages	High	Sprint-1
Customer (Web user)	Upload Images	USN-6	As a user, I can upload theimage that required for finding whether liver disease is there are not.	Can get resultbased on the information provided.	Medium	Sprint-3
	Enter data	USN-7	Asa user, I can enter the required data from the scanned report.	Canget result based on the information provided.	High	Sprint-3
	Report	USN-8	As a user, I can generate thereport in PDFformat.	Result can be generated in PDF format in user login.	Low	Sprint-4

	Search	USN-9	As a user, I can search for thespecialist andbest hospital in that respective field.	I can receive information on various doctors and hospitals.	Low	Sprint-4
Administrator	Analyse	USN-10	As an admin, I will analyse thegiven data.	I can analyse thegiven data.	High	Sprint-2

6.PROJECT PLANNING AND SCHEDULING

6.1 Sprint Planning and Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	10	High	Anupama Jeyashree S
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application.	5	High	Lavanya M
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email & password.	5	High	Parvathi Priya Nandana K M
Sprint -2	Dashboard	USN-4	As a user, I must enter my details.	15	Medium	Anithashree V
Sprint -2	Dashboard	USN-5	As a user, I can navigate through different pages using the	5	High	Anithashree V

			dashboard.			
Sprint - 3	Upload Images	USN-6	As a user, I can upload the image that required for finding whether liver disease is there are not.	5	Medium	Lavanya M
Sprint-3	Enter Data	USN-7	As a user, I can enter the required data from the scanned report.	15	High	Anupama Jeyashree S
Sprint - 4	Display Result	USN-8	As a user, I can view the result.	15	High	Parvathi Priya Nandana K M

6.2 Sprint Delivery Schedule

Sprint	**Total Story Point**	**Duration**	**Sprint Start Date**	**Sprint End Date (Planned)**	**Story Points Completed (as on Planned End Date)**	**Sprint Release Date (Actual)**
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

6.3 Reports from JIRA

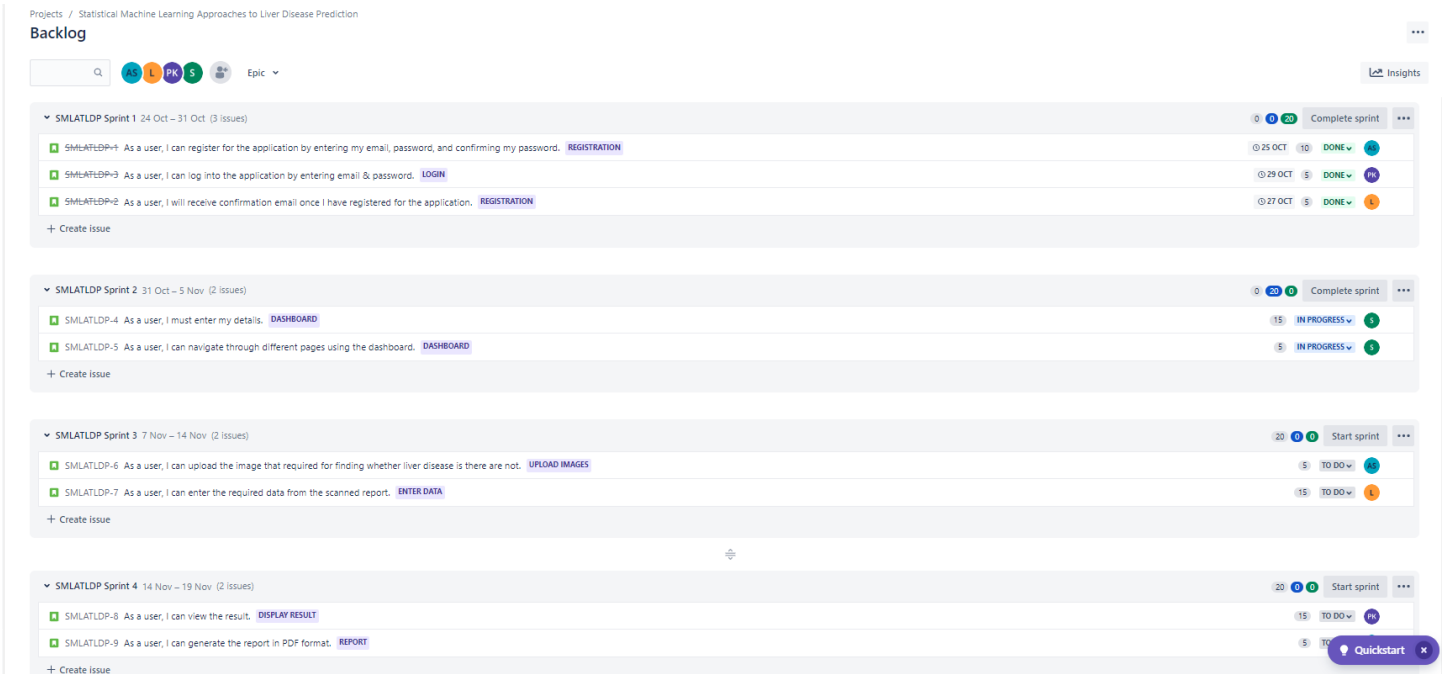


FIG 6.3.1 Backlog in JIRA

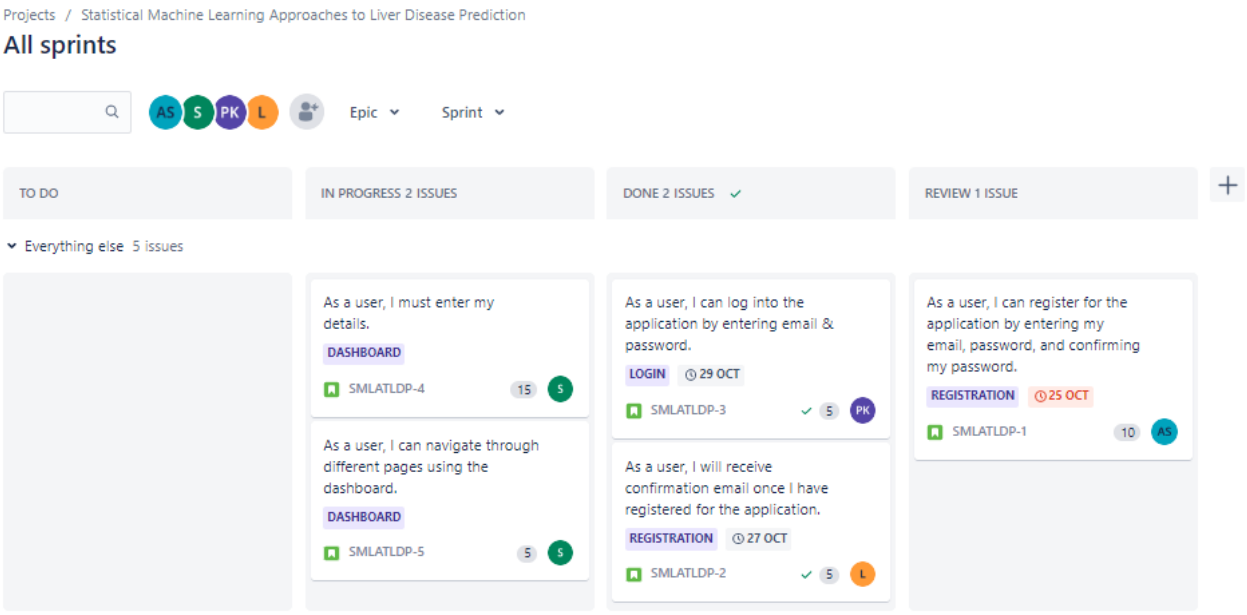


FIG 6.3.2 Sprint in JIRA

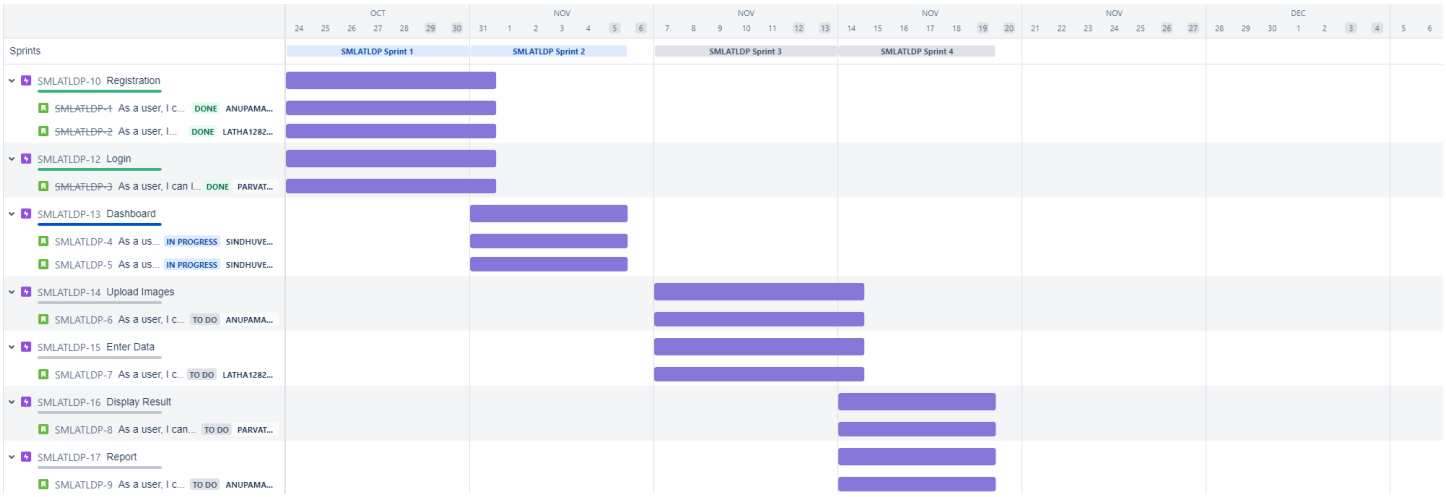


FIG 6.3.3 Progress Tracking in JIRA

7. CODING & SOLUTIONING

7.1 Feature

The application being proposed here uses concept of machine learning, and the models are first trained, then tested. The user can interact with the system through the developed web application. At first, the system asks you to register for the application by entering email id, password and confirming my password. One we have register, we will get a confirmation email. Then log in to the application by entering the email and password. A Dashboard will be opened, the registered user can access the dashboard and can also retrieve information anywhere.

The user can navigate through different pages using the dashboard. For the prediction process, the user should upload the image that is required for finding whether liver disease is there are not. The main feature is the result of the tested data will be generated as report in the form of PDF and stored in the user login and also it willautomatically download to the user system.

PREDICTION

app.py

```
from flask import Flask, render_template, request
from flask_cors import CORS
import joblib
import requests

# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "Ajx7BZ64A-kNZlcGlAyn4Dr-RYfooJomuf0V44dWXDDZ"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
 API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]

header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}

app = Flask(__name__)

@app.errorhandler(404)
def page_not_found(e):
    return render_template('404.html'), 404

@app.route('/')
def index():
```

```
    return render_template('index.html')

@app.route('/prediction_form')
def predict1():
    return render_template('predict.html')

@app.route('/predict', methods=['POST'])
def predict():
    age = float(request.form['Age'])
    gender = float(request.form['Gender'])
    tb = float(request.form['Total_Bilirubin'])
    db = float(request.form['Direct_Bilirubin'])
    ap = float(request.form['Alkaline_phosphatase'])
    aa1 = float(request.form['Alamine_Aminotransferase'])
    aa2 = float(request.form['Aspartate_Aminotranferase'])
    tp = float(request.form['Total_Proteins'])
    al = float(request.form['Albumin'])
    agr = float(request.form['Albumin_and_Globulin_Ratio'])

    # converting data into float
    X = [[age,gender,tb,db,ap,aa1,aa2,tp,al,agr]]

    # NOTE: manually define and pass the array(s) of values to be scored in the next line
    payload_scoring = {"input_data": [{"field": [['age','gender','tb','db','ap','aa1','aa2','tp','al','agr']], "values": X }]}

    response_scoring = requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/06acb9ea-228d-4770-
8841-afb3a9cedc6c/predictions?version=2022-11-16', json=payload_scoring,
    headers={'Authorization': 'Bearer ' + mltoken})
    print(response_scoring.json())
    print(response_scoring)
    predictions = response_scoring.json()
    predict = predictions['predictions'][0]['values'][0][0]
    print(f"Final prediction : {predict}")
    # prediction = model.predict(data)[0]

    return render_template('result.html',result = predict)

if __name__ == '__main__':
    app.run(debug=True)


# SVC from version 1.0.2 when using version 1.1.3.
```

```
# 2 - have disease
# 1 - Not have disease
```

Index.html

```
<!DOCTYPE html>
<html>
<head>
 <title></title>
 <link rel="stylesheet" type="text/css" href="../static/css/style.css">
 <link href="https://fonts.googleapis.com/css?family=Josefin+Sans&display=swap" rel="stylesheet">
<script src="https://kit.fontawesome.com/66cfa4dafa.js" crossorigin="anonymous"></script>
</head>
<body>
<header>
 <div class="mainheader">
  <div class="logo">
   <img src="../static/images/logo.png">
  </div>

  <nav>
   <a href="http://127.0.0.1:5000">home</a>
   <a href="#hlo">About</a>
  </nav>

  <div class="menubtn">
   <a href="/prediction_form"><button> Predict</button></a>
  </div>
 </div>




 <main>
  <section class="left-sec">
   <h2> We Are Here For Your Easy Diagnosis</h2>
   <h1>  We The Best Predictors</h1>
   <p>We are here to make your care easy. Feel free to use it.</p>
   <a href="/prediction_form"><button>Predict</button></a>
```

```
 </section>

 <section class="right-sec">
  <figure>
   <img src="../static/images/Home_image.png">
  </figure>
 </section>
</main>
</header>
<section class="about"  >
      <div class="main-about">
                <img src="../static/images/about.jpg" alt="About">
   <div class="about-text" id="hlo">
      <h1>Liver Disease </h1>
      <h5>Introduction </h5>
      <p>The liver plays an important role in many bodily functions from protein production and blood clotting to cholesterol, glucose (sugar), and iron metabolism. Many diseases and conditions can affect the liver, for example, certain drugs like excessive amounts of acetaminophen, and acetaminophen combination medications like Vicodin and Norco, as well as statins, cirrhosis, alcohol abuse, hepatitis A, B, C, D, and E, infectious mononucleosis (Epstein Barr virus), nonalcoholic fatty liver disease (NASH), and iron overload (hemochromatosis). Liver disease is a broad term that covers all the potential problems that cause the liver to fail to perform its designated functions. Usually, more than 75% or three quarters of liver tissue needs to be affected before a decrease in function occurs. This liver disease predictor helps you to predict liver disease based on the following datas.</p>

   </div>
        </div>

   </section>
<section>
<div class="service">
<div class="title">
<h2>Information About Different Liver Enzymes</h2>
</div>
<div class="box">
<div class="card">
<i class="fa-solid fa-circle-info"></i>
<h5>Total bilirubin</h5>
<div class="pra">
<p>This is a blood test that measures the amount of a substance called bilirubin. This test is used to find out how well your liver is working. It is often given as part of a panel of tests that measure liver function. A small amount of bilirubin in your blood is normal, but a high level may be a sign of liver disease. The liver makes bile to help you
```

```
digest food, and bile contains bilirubin. Most bilirubin comes from the body's normal process of breaking down old red blood cells. A healthy liver can normally get rid of bilirubin. But when you have liver problems, bilirubin can build up in your body to unhealthy levels.</p>
<p style="text-align:center;">
<a class="button" >Range: .22 - 1.0 mg/dl</a>
</p>
</div>
</div>
<div class="box">
<div class="card">
<i class="fa-solid fa-circle-info"></i>
<h5>Direct bilirubin</h5>
<div class="pra">
<p>The diagnosis is narrowed down further by evaluating the levels of direct bilirubin. If direct (conjugated) bilirubin is normal, then the problem is an excess of unconjugated bilirubin (indirect bilirubin), and the location of the problem is upstream of bilirubin conjugation in the liver. Hemolysis, or internal haemorrhage can be suspected. If direct bilirubin is elevated, then the liver is conjugating bilirubin normally, but is not able to excrete it. Bile duct obstruction by gallstones, hepatitis, cirrhosis or cancer should be suspected.</p>
<p style="text-align:center;">
<a class="button" >Range: 0.0 - 0.2 mg/dl</a>
</p>
</div>
</div>
<div class="box">
<div class="card">
<i class="fa-solid fa-circle-info"></i>
<h5>Alkaline phosphatase</h5>
<div class="pra">
<p>Alkaline phosphatase (ALP) is an enzyme in the cells lining the biliary ducts of the liver. ALP levels in plasma rise with large bile duct obstruction, intrahepatic cholestasis, or infiltrative diseases of the liver. ALP is also present in bone and placental tissue, so it is higher in growing children (as their bones are being remodelled) and elderly patients with Paget's disease. In the third trimester of pregnancy, ALP is about two to three times higher. Biliary tract disease produces relatively greater increases in ALP than increases in ALT, AST, or LD. ALP is associated with the plasma membrane of hepatocytes adjacent to the biliary canaliculus. Obstruction or inflammation of the biliary tract results in an increased concentration of the ALP in the circulation. Similar to ALT and AST, ALP is not specific for biliary tract disease. ALP is released by osteoblasts, the ileum, and the placenta. </p>
<p style="text-align:center;">
<a class="button" >Range: 110 - 310 U/L</a>
</p>
</div>
</div>
```

```
</div>
</div>
</section>
<section>
<div class="service1">
<div class="box">
<div class="card">
<i class="fa-solid fa-circle-info"></i>
<h5>Alamine Aminotransferase</h5>
<div class="pra">
<p>ALT is normally found inside liver cells. However, when your liver is damaged or inflamed, ALT can be released into your bloodstream. This causes serum ALT levels to rise. Measuring the level of ALT in a person’s blood can help doctors evaluate liver function or determine the underlying cause of a liver problem. The ALT test is often part of an initial screening for liver disease. The ALT test is usually used to determine whether someone has liver injury or failure. Your doctor may order an ALT test if you’re having symptoms of liver disease, including: • jaundice, which is yellowing of your eyes or skin • dark urine • nausea • vomiting • pain in the right upper quadrant of your abdomen.</p>
<p style="text-align:center;">
<a class="button" >Range: 5 - 45 U/L</a>
</p>
</div>
</div>
<div class="box">
<div class="card">
<i class="fa-solid fa-circle-info"></i>
<h5>Aspartate Aminotranferase</h5>
<div class="pra">
<p>Aspartate aminotransferase (AST) test is a blood test that checks for liver damage. Your doctor might order this test to find out if you have liver disease and to monitor your treatment. Your liver is an organ that has many important jobs. It makes a fluid called bile that helps your body digest food. It also removes waste products and other toxins from your blood. It produces proteins, as well as substances that help your blood clot. Alcohol or drug use and diseases such as  hepatitis can damage your liver and keep it from doing these jobs.</p>
<p style="text-align:center;">
<a class="button" >Range: 5 - 40 U/L</a>
</p>
</div>
</div>
<div class="box">
<div class="card">
<i class="fa-solid fa-circle-info"></i>
<h5>Albumin</h5>
```

```
<div class="pra">
<p>Albumin is a protein made specifically by the liver, and can be measured cheaply and easily. It is the main constituent of total protein (the remaining constituents are primarily globulins). Albumin levels are decreased in chronic liver disease, such as cirrhosis. It is also decreased in nephrotic syndrome, where it is lost through the urine. The consequence of low albumin can be edema since the intravascular oncotic pressure becomes lower than the extravascular space. An alternative to albumin measurement is prealbumin, which is better at detecting acute changes (half-life of albumin and prealbumin is about 2 weeks and about 2 days, respectively). </p>
<p style="text-align:center;">
<a class="button" >Range: 3.5 - 5 gm/dl</a>
</p>
</div>
</div>
</div>
</div>
</section>
<section>
<div class="service1">
<div class="box">
<div class="card">
<i class="fa-solid fa-circle-info"></i>
<h5>Total Proteins</h5>
<div class="pra">
<p>The total protein test measures the total amount albumin and globulin in your body. It's used as part of your routine health checkup. Having too many or too few proteins can lead to unexpected weight loss, fatigue, or inflammatory disease.Proteins serve as building blocks for many organs, hormones, and enzymes. Proteins are essential for overall health, which is why routine health checkups often include a total protein test.A total protein test measures the total number of proteins present in body fluid. The test examines protein in either urine or the liquid portion of the blood, which medial professionals call the serum. The total protein test can help diagnose liver and kidney diseases, along with other conditions.
</p>
<p style="text-align:center;">
<a class="button" >Range: 5 - 45 U/L</a>
</p>
</div>
</div>
<div class="box">
<div class="card">
<i class="fa-solid fa-circle-info"></i>
<h5>Albumin and Globulin Ratio</h5>
<div class="pra">
<p>The Albumin to Globulin ratio (A:G) is the ratio of albumin present in serum in relation to the amount of globulin.
```

```
The ratio can be interpreted only in light of the total protein concentration. The normal ratio in most species approximates 1:1.Serum proteins are primarily albumin (50–60%, produced by the liver), but also include globulins and other proteins. Serum proteins maintain water balance in the blood through osmotic pressure, transport blood components and nutritional elements, help the immune system and help with coagulation. Although albumin is made exclusively in the liver, globulins are produced in many sites throughout the body. Thus, whether total protein is normal, elevated, or low, a decrease in the albumin:globulin (A/G ratio) often indicates the presence of impaired liver function.</p>
<p style="text-align:center;">
<a class="button" >Range: 1.7-2.2</a>
</p>
</div>
</div>
</div>
</div>
</section>
</body>
</html>
```

Predict.html

```
<!DOCTYPE html>
<html >
<head>
  <meta charset="UTF-8">
  <title>Prediction</title>


  <link rel='stylesheet prefetch' href='https://cdnjs.cloudflare.com/ajax/libs/ionicons/2.0.1/css/ionicons.css'>
<link rel='stylesheet prefetch' href='https://maxcdn.bootstrapcdn.com/font-awesome/4.6.3/css/font-awesome.min.css'>
<script src="https://kit.fontawesome.com/66cfa4dafa.js" crossorigin="anonymous"></script>
      <link rel="stylesheet" href="../static/css/style1.css">


</head>

<body>

 <section>
```

```
 <header>
 <div class="mainheader">
  <div class="logo">
   <img src="../static/images/logo.png">
  </div>

  <nav>
   <a href="http://127.0.0.1:5000">home</a>
   <a href="http://127.0.0.1:5000#hlo" >About</a>
  </nav>
 </div>
</section>

 <div class="signupSection">
 <div class="info">
   <h2>Mission to Diagnose Earlier</h2>
  <div class="logo">
  <img src="../static/images/logo_2.svg">
 </div>
   <p>The Future Is Here</p>
 </div>
 <form action="/predict" method="POST" class="signupForm" name="signupform">
   <h2>Liver Disease Prediction</h2>
   <ul class="noBullet">
     <li >
       <label for="Age"></label>
       <input type="number" class="inputFields" id="Age" name="Age" placeholder="Age" value="" oninput="return
userNameValidation(this.value)" required/>

       <label for="Gender"></label>
       <input type="number" class="inputFields" id="Gender" name="Gender" placeholder="Enter 1 for male, 0 for
female" value="" oninput="return passwordValidation(this.value)" required/>
     </li>
     <li>
       <label for="Direct Bilirubin"></label>
       <input type="number" step=0.1 class="inputFields" id="Total_Bilirubin" name="Total_Bilirubin"
placeholder="Total Bilirubin" value="" required/>

       <label for="Total Bilirubin"></label>
       <input type="number" step=0.01 class="inputFields" id="Direct_Bilirubin" name="Direct_Bilirubin"
placeholder="Direct Bilirubin" value="" required/>
```

```
   </li>
    <li>
      <label for="Alkaline phosphatase"></label>
      <input type="number" class="inputFields" id="Alkaline_phosphatase" name="Alkaline_phosphatase"
placeholder="Alkaline Phosphatase" value="" required/>

      <label for="Alamine Aminotransferase"></label>
      <input type="number" class="inputFields" id="Alamine_Aminotransferase" name="Alamine_Aminotransferase"
placeholder="Alamine Aminotransferase" value="" required/>
     </li>
     <li>
      <label for="Aspartate Aminotranferase"></label>
      <input type="number" class="inputFields" id="Aspartate_Aminotranferase" name="Aspartate_Aminotranferase"
placeholder="Aspartate Aminotranferase" value="" required/>

      <label for="Albumin"></label>
      <input type="number" step=0.1 class="inputFields" id="Albumin" name="Albumin" placeholder="Albumin"
value="" required/>
    </li>
   <li>
      <label for="Total Proteins"></label>
      <input type="number" step =0.1 class="inputFields" id="Total_Proteins" name="Total_Proteins"
placeholder="Total Proteins" value="" required/>

      <label for="Albumin and Globulin Ratio"></label>
      <input type="number" step=0.1 class="inputFields" id="Albumin_and_Globulin_Ratio"
name="Albumin_and_Globulin_Ratio" placeholder="Albumin and Globulin Ratio" value="" required/>
    </li>

    <li id="center-btn">
      <input type="submit" id="predict-btn" name="predict" alt="Predict" value="Predict">
    </li>
   </ul>
  </form>
</div>
 <script src="../static/js/index.js"></script>

</body>

</html>
```

Result.html

```
<!DOCTYPE html>
<html >
<head>
  <meta charset="UTF-8">
  <title>Prediction</title>


  <link rel='stylesheet prefetch' href='https://cdnjs.cloudflare.com/ajax/libs/ionicons/2.0.1/css/ionicons.css'>
<link rel='stylesheet prefetch' href='https://maxcdn.bootstrapcdn.com/font-awesome/4.6.3/css/font-awesome.min.css'>
<script src="https://kit.fontawesome.com/66cfa4dafa.js" crossorigin="anonymous"></script>
      <link rel="stylesheet" href="../static/css/style2.css">


</head>

<body>

 <section>
  <header>
 <div class="mainheader">
  <div class="logo">
   <img src="../static/images/logo.png">
  </div>

  <nav>
   <a href="http://127.0.0.1:5000">home</a>
   <a href="http://127.0.0.1:5000#hlo">About</a>
  </nav>
 <div class="menubtn">
  <a href="/prediction_form"> <button> Predict</button></a>
  </div>
 </div>
</header>
</section>


<section>
  <div class="healthy">
    {% if result == 1 %}
```

```
    <img src="{{url_for('static',filename='Images/unhealthy.jpeg')}}" alt="damage" >
    {% elif result == 2 %}
    <img src="{{url_for('static',filename='Images/healthy.jpeg')}}" alt="healthy">
    {% endif %}
   </div>
</section>

</body>

</html>
```

8. TESTING

Test Scenarios

S. No	Testing Scenarios
1	Verify user is able to see home page
2	Verify user is able to view the about section of the home page
3	Verify user is able to navigate to predict the liver disease page
4	Verify user is able to come back to home page from result and prediction page
5	Verify predict page elements
	Input Fields
1	Verify user is able to enter float values in input field of the prediction
2	Verify user is able to enter number in some of the input field in prediction page
3	Verify user is able to predict only after entering all the fields

8.1 Test Cases

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BUG ID	Executed By
HomePage_TC_001	UI	Home Page	Verify user is able to see all the buttons and all the contents of the page		1.Enter URL and click go 2.Explore the whole page to verify that all the fonts are in correct size and readable	http://127.0.0.1:5000/	All the elements and contents should be readable font color and size and with uniformity	Working as expected	Pass				Lavanya M
HomePage_TC_OO2	Functional	Home Page	Verify user is able to see the about section of the page when clicked about on the home page		1.Enter URL and click go 2.Click on about hyperlink on the page 3.Verify it moves down to about section of the page	http://127.0.0.1:5000/	About section of the home page should display	Working as expected	Pass				Lavanya M
HomePage_TC_OO3	Functional	Home Page	Verify the prediction form page is displayed when the predict button is clicked		1.Enter URL and click go 2.Click on predict button on the home page 3.Verify the prediction page is displayed.	http://127.0.0.1:5000/	Application should show prediction page to enter details	Working as expected	Pass				Lavanya M
PredictionPage_TC_ OO1	Functional	Prediction page	Verify user is able to enter values in the input fields in both float and integer values.		1.Enter URL(http://127.0.0.1:5000) and click go 2.Click on Predict button 3.Enter the details in the input field box in float and integer wherever necessary 4.click predict button.	Age: 17 Gender: 0 Total Bilirubin: 0.9 Direct Bilirubin: 0.3 Alkaline Phosphatase: 202 Alamine Aminotransferase: 22 Asparate Aminotransferase: 19 Albumin: 7.4 Total Proteins: 4.1 Albumin and Globulin Ratio: 1.2	User should navigate to result page	Working as expected	Pass				Lavanya M
PredictionPage_TC_ OO2	Functional	Prediction page	Verify user can view the predicted result only after entering all the required field data		1.Enter URL(http://127.0.0.1:5000) and click go 2.Click on Predict button 3.Enter the details in the input field box in float and integer wherever necessary 4.click predict button.	Age: 17 Gender: 0 Total Bilirubin: 0.9 Direct Bilirubin: 0.3 Alkaline Phosphatase: 202 Alamine Aminotransferase: 22 Asparate Aminotransferase: 19 Albumin: - Total Proteins: 4.1 Albumin and Globulin Ratio: 1.2	Application should show "Enter all the required fields"(Albumin)	Working as expected	Pass				Lavanya M

PredictionPage_TC_OO3		Login page	Verify user is able to click the Home and predict again button		1.Enter URL(http://127.0.0.1:5000) and click go 2.Click on Predict button 3.Enter the details in the input field box in float and integer wherever necessary 4.click predict button. 5. View the result 6. Click the predict button/home button		Application should move to Prediction page or home page respectively.	Working as expected	Pass				Lavanya M

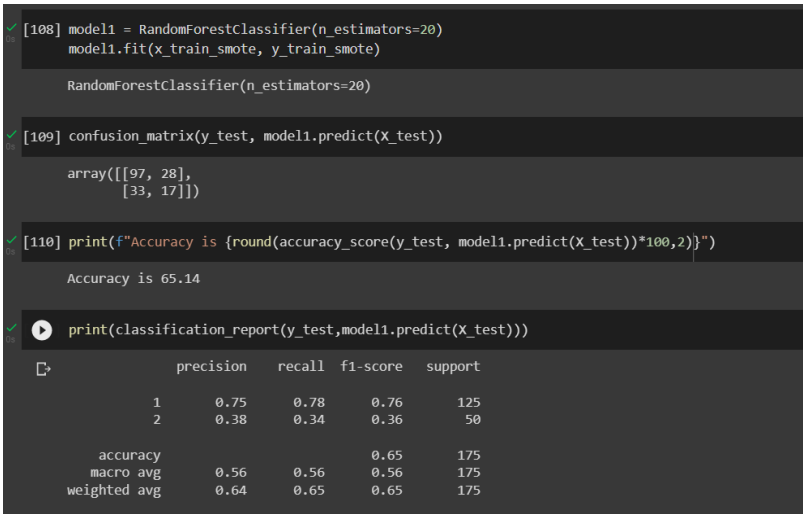
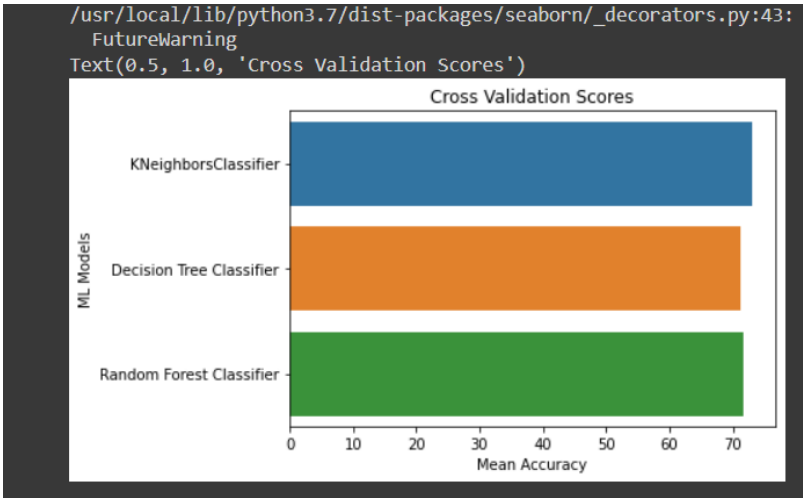
8.2 User Acceptance Testing

This report shows the numberof test cases that have passed, failed,and untested.

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	2	0	0	2
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	1	0	0	1
Exception Reporting	1	0	0	1
Final Report Output	3	0	0	3
Version Control	1	0	0	1

9. RESULTS

9.1 Performance Metrics

S. No.	Parameter	Values	Screenshot
1.	Metrics	**Classification Model:** Confusion Matrix -, Accuracy Score- & Classification Report -	
2.	Tune the Model	Hyperparameter Tuning - Validation Method -	

```
[108] model1 = RandomForestClassifier(n_estimators=20)
      model1.fit(x_train_smote, y_train_smote)

      RandomForestClassifier(n_estimators=20)

[109] confusion_matrix(y_test, model1.predict(X_test))

      array([[97, 28],
             [33, 17]])

[110] print(f"Accuracy is {round(accuracy_score(y_test, model1.predict(X_test))*100,2)}")

      Accuracy is 65.14

      print(classification_report(y_test,model1.predict(X_test)))

                   precision    recall  f1-score   support

                1       0.75      0.78      0.76       125
                2       0.38      0.34      0.36        50

         accuracy                           0.65       175
        macro avg       0.56      0.56      0.56       175
     weighted avg       0.64      0.65      0.65       175
```

FIG 9.1.1 Metrics of Classification Model

FIG 9.1.2 Hyperparameter tuning

10. ADVANTAGES AND DISADVANTAGES

Advantages

- This helps in early diagnosis of liver disease.
- It makes the process simple and easier.
- We are able to monitor the patient in an effective manner.
- We can also help patients who are not in a condition to directly consult a doctor.

Disadvantages

- There are inaccuracies which cannot be avoided.
- For people who have difficulty in accessing internet or those who do not have electronic gadgets , this is not a feasible solution.

11. CONCLUSION

The main aim of this project is to identify whether a user has liver disease or not based upon the data entered as input. Based on the inputs entered by the user, a result is displayed**.** If the user has been diagnosed with any form of liver disease, then it is highly recommended that they seek medical attention immediately. With the help of this application, the user will be able to save both time and money and the user can diagnose the disease without visiting any labs or hospitals, and this is only applicable to mild diseases. The proposed model achieved an accuracy of 67%.

12. FUTURE SCOPE

- In biomedical science, accuracy and speed are two important factors that should be considered chiefly in dealing with any disease.
- In this regard, Machine Learning techniques can be of great help to physicians. With advances, several machines have entered in our lives.
- One of the most famous areas where computers as the mostly used machines can be helpful is knowledge extraction with the help of a machine (machine learning).

13. APPENDIX

Source Code

https://github.com/IBM-EPBL/IBM-Project-26211-1660021283/tree/main/Final%20Deliverables/Deployment

Github Link

https://github.com/IBM-EPBL/IBM-Project-26211-1660021283

Project Demo Link

https://drive.google.com/file/d/1wOmXct0U9o3tySK0xYf15OvdDishXR_y/view?usp=sharing