

STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION

PROJECT REPORT

**SUBMITTED BY
TEAM ID:PNT2022TMID36186**

VINOTH.J	110519104032
JENIFER.A	110519104008
JAYAPRIYA.R	110519104007
AARTHI.R	110519104001
SESHAN.M	110519104023

**In Partial fulfilment for the Award of Award of the Degree
OF
BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE AND ENGINEERING**



**GOJAN SCHOL OF BUSINESS AND TECHNOLOGY
REDHILLS.**

**ANNA UNIVERSITY : CHENNAI - 600 025
NOVEMBER 2022**

TABLE OF CONTENTS

1. INTRODUCTION

- a. Project Overview
- b. Purpose

2. LITERATURE SURVEY

- a. Existing problem
- b. References
- c. Problem Statement Definition

3. IDEATION & PROPOSED SOLUTION

- a. Empathy Map Canvas
- b. Ideation & Brainstorming
- c. Proposed Solution
- d. Problem Solution fit

4. REQUIREMENT ANALYSIS

- a. Functional requirement
- b. Non-Functional requirements

5. PROJECT DESIGN

- a. Data Flow Diagrams
- b. Solution & Technical Architecture
- c. User Stories

6. PROJECT PLANNING & SCHEDULING

- a. Sprint Planning & Estimation
- b. Sprint Delivery Schedule
- c. Reports from JIRA

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

- a. Feature 1
- b. Feature 2
- c. Database Schema (if Applicable)

8. TESTING

- a. Test Cases
- b. User Acceptance Testing

9. RESULTS

- a. Performance Metrics

10. ADVANTAGES & DISADVANTAGES

11. CONCLUSION

12. FUTURE SCOPE

13. APPENDIX

Source Code

GitHub & Project Demo Link

1. INTRODUCTION

1.1 PROJECT OVERVIEW

Liver diseases avert the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises. Early prediction of liver disease using classification algorithms is an efficacious task that can help the doctors to diagnose the disease within a short duration of time. Discovering the existence of liver disease at an early stage is a complex task for the doctors. The main objective of this project is to analyze the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease.

1.2 PURPOSE

This Project examines data from liver patients concentrating on relationships between a key list of liver enzymes, proteins, age and gender using them to try and predict the likeliness of liver disease. Here we are building a model by applying various machine learning algorithms find the best accurate model. And integrate to flask based web application. User can predict the disease by entering parameters in the web application.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

Liver diseases have produced a big data such as metabolomics analyses, electronic health records, and report including patient medical information, and disorders. However, these data must be analyzed and integrated if they are to produce models about physiological mechanisms of pathogenesis. We use machine learning based on classifier for big datasets in the fields of liver to Predict and therapeutic discovery. A dataset was developed with twenty three attributes that include the records of 7000 patients in which 5295 patients were male and rests were female. Support Vector Machine (SVM), Boosted C5.0, and Naive Bayes (NB), data mining techniques are used with the proposed model for the prediction of liver diseases. The performance of these classifier techniques are evaluated with accuracy, sensitivity, specificity.

2.2 REFERENCES:

1. Thomas G. Cotter, MARY RENILA, Nonalcoholic Fatty Liver Disease 2020: The State of the Disease, Gastroenterology, Volume 158, Issue 7, 2020, Pages 1851-1864, ISSN 0016-5085, <https://doi.org/10.1053/j.gastro.2020.01.052>.
- 2.K.THIRUNAVUKARASU, A. S. Singh, M. IRFAN and A. Chowdhury, "Prediction of Liver Disease using Classification Algorithms," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-3,DOI: 10.1109/CCAA.2018.8777 655.
3. M. A. KUZHIPALLIL, C. Joseph and A.KANNAN, "Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 778-782, DOI: 10.1109/ICACCS48705.2020.9074368
4. V. J. GOGI and V. M.N., "Prognosis of Liver Disease: Using Machine Learning Algorithms," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), Bhubaneswar, India, 2018, pp. 875-879, DOI: 10.1109/ICRIEECE44171.2018.9008482.
5. N. Li et al., "Machine Learning Assessment for Severity of Liver Fibrosis for Chronic HBV Based on Physical Layer With Serum Markers," in IEEE Access, vol. 7, pp. 124351-124365, 2019, DOI: 10.1109/ACCESS.2019.2923688.

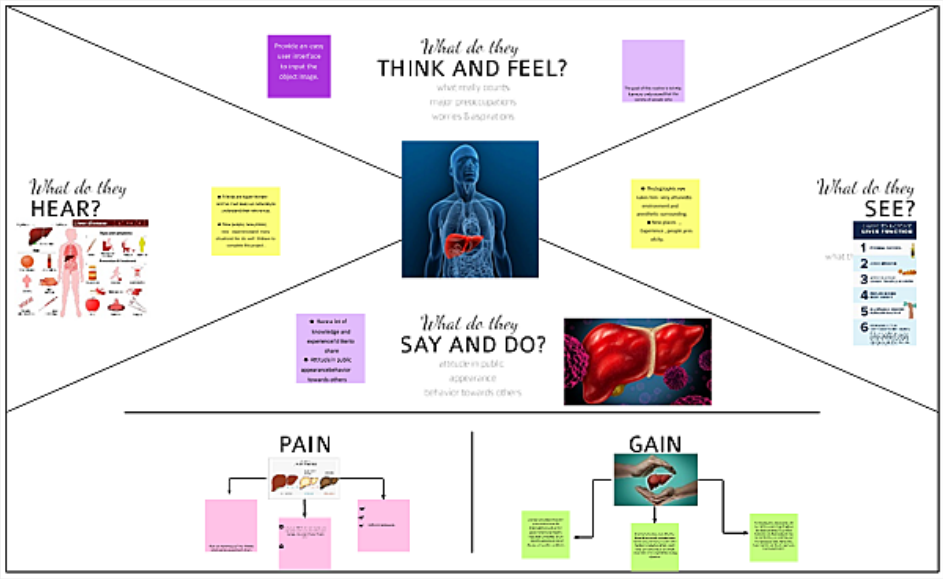
2.3 PROBLEM STATEMENT DEFINITION



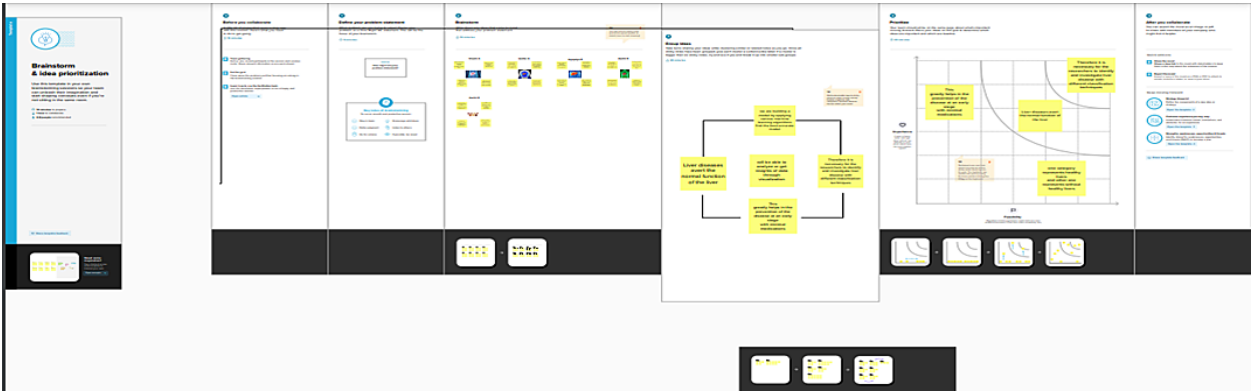
PROBLEM STATEMENT(PS)	I AM (CUSTOMER)	I'M TRYING TO	BUT	BECAUSE	WHICH MAKES ME FEEL
PS-1	Predictor	Predict the liver disease	It takes a long time	It is difficult to cure disease	Frustration
PS-2	Patient	Lead a healthy life	I am unable to start a healthy life for my future	I am addicted to some drugs, etc...	Risk

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION & BRAINSTORMING



3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
•	Problem Statement (Problem to be solved)	1)Liver diseases avert the normal function of the liver. 2)Mainly due to the large amount of alcohol consumption liver disease arises.
•	Idea / Solution description	Early prediction of liver disease using classification algorithms is an efficacious task that can help the doctors to diagnose the disease within a short duration of time.
•	Novelty / Uniqueness	Discovering the existence of liver disease at an early stage is a complex task for the doctors.
•	Social Impact	1)Liver disease can be inherited (genetic). 2)Liver problems can also be caused by a variety of factors that damage the liver, such as viruses, alcohol use and obesity.
•	Business Model (Revenue Model)	1)This can be implemented as an essential detection method in every country. 2)Accurate detection and analysis can encourage the increase in financial benefit.
•	Scalability of the Solution	1) Accurate predictions and extensive use. 2) Availability.

3.4 PROBLEM SOLUTION FIT

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS Who is your customer? i.e. working parents of 0-5 yrs kids Our customers are the patients who are suffering from liver disease. Especially occurs due to the large amount of alcohol consumption. Currently, the liver related diseases are identified by analyzing liver function blood test reports and scan reports. It takes more time as well as expensive. It is not sure that the accuracy of diagnosing the disease is best in existing solutions.	6. CUSTOMER CONSTRAINTS CC What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices. -----Should have smartphones -----Should have internet access	5. AVAILABLE SOLUTIONS AS Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking In earlier days, there is a traditional approach to diagnosing liver disease are by using algorithms like --Naive Bayes Classifier --Support Vector Machines --Back Propagation Neural Network --Decision tree --Random tree and so on. But they are failed due to uncertainty in accuracy.	Explore AS, differentiate
	2. JOBS-TO-BE-DONE / PROBLEMS J&P Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides. The problems which we have encountered are Accuracy --- The model should acquire required accuracy because it involves the risk of life of human beings. Identify --- There are different kinds of liver disease and so our model should be able to predict all kinds of liver disease. Risk Involved --- The model should be able to predict the level of risk that the patient currently have due to the diagnosed disease.	9. PROBLEM ROOT CAUSE RC What is the real reason that this problem exists? What is the back story behind the need to do this job? i.e. customers have to do it because of the change in regulations. The root cause of the problems are --- Acquiring proper dataset is difficult. --- Parameters used for the training and testing the dataset should be able to predict any kind of liver disease and risks involved if the person is diagnosed with the particular disease. --- The model may require more real-time data to improve its accuracy and so there may be uncertainty in the predicted result at the start of the app released.	7. BEHAVIOUR BE What does your customer do to address the problem and get the job done? i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace) ----- People may stop using the application if the predicted results are not appropriate. ----- People may also try to use applications which has better response speed. ----- They avoid to use the predictors if it is no user-friendly.	
Focus on J&P, tap into BE, understand RC	3. TRIGGERS TR ---People wants to make their life easier, use it anywhere and anytime. ---Now-a-days web application is the one which is easily accessible as they don't like to download lots of app in their mobile.	10. YOUR SOLUTION SL Our solution to solve this problem is to develop ---An application which is accessible from anywhere at anytime using their mobile/laptop/tablet. ---Try to develop the application with more accuracy. ---Try to develop the application with as many as features possible to give more benefits to the consumer.	8. CHANNELS OF BEHAVIOUR CH 8.1 ONLINE ---People may be able to access the application in the browser from anywhere at anytime. ---Advertise about the application with influencers to promote the application. 8.2 OFFLINE --- Word of mouth among consumers (especially doctors).	Extract online & offline CH of BE
	4. EMOTIONS: BEFORE / AFTER EM ---People find it difficult to trust the predicted results. So, our goal is to work on accuracy and change it. ---People will feel easier to access the application and can be able to diagnose the liver disease in their house itself and can ensure security of their records.			

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User input	Get necessary details for prediction
FR-4	Data Processing	Data cleaning, Data scaling, Feature selection
FR-5	Prediction	Predicting whether the user has liver disease or not

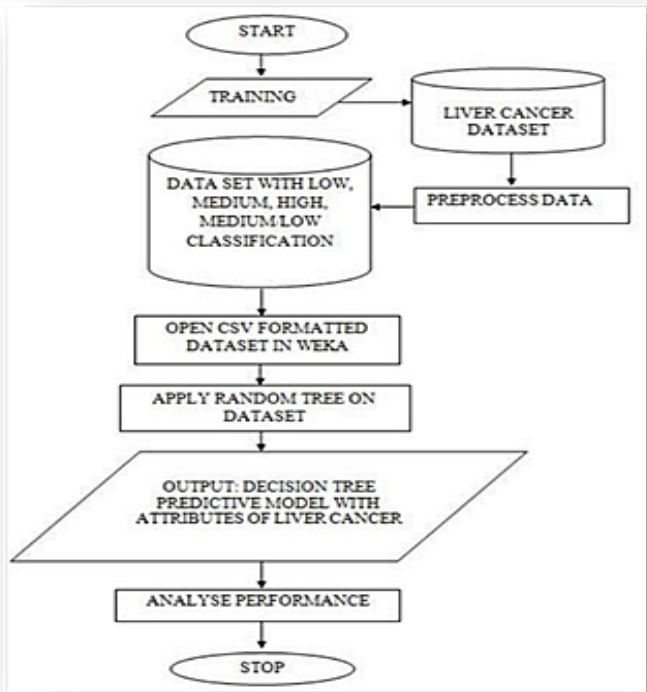
4.2 NON FUNCTIONAL REQUIREMENT

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	To check whether the patient has liver disease or not
NFR-2	Security	Implement necessary techniques to provide security to the user data
NFR-3	Reliability	Make ensure that the model is reliable
NFR-4	Performance	Use efficient ML techniques for better accuracy
NFR-5	Availability	By having few basic data set of people we can predict the disease.
NFR-6	Scalability	Predicts various types of liver disease

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	To check whether the patient has liver disease or not
NFR-2	Security	Implement necessary techniques to provide security to the user data
NFR-3	Reliability	Make ensure that the model is reliable
NFR-4	Performance	Use efficient ML techniques for better accuracy
NFR-5	Availability	By having few basic data set of people we can predict the disease.
NFR-6	Scalability	Predicts various types of liver disease

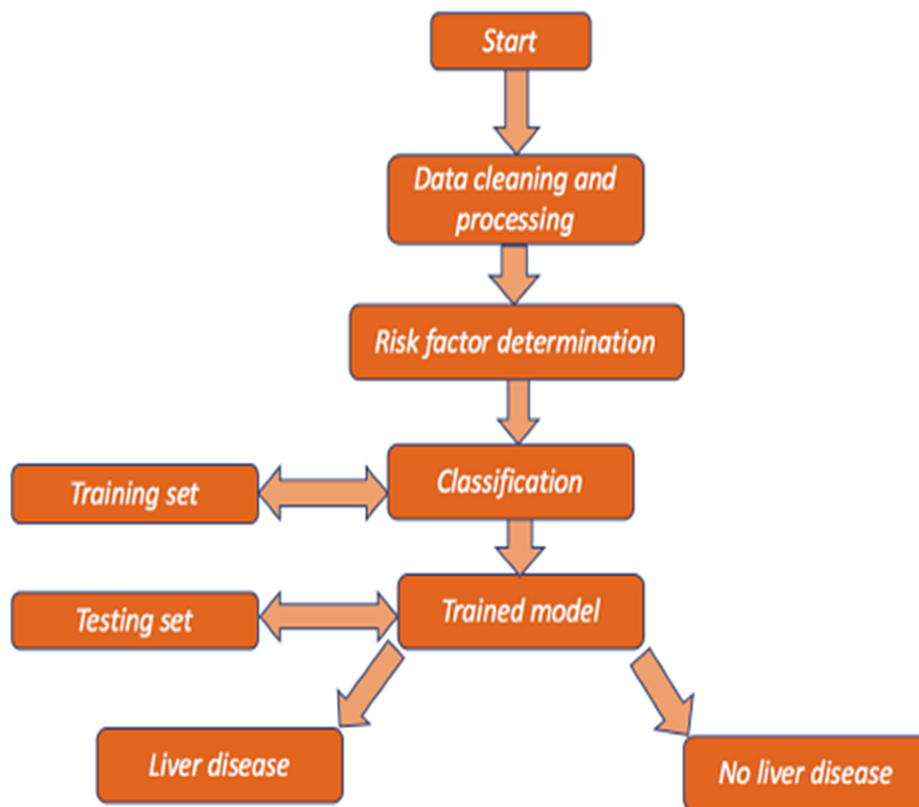
5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAM

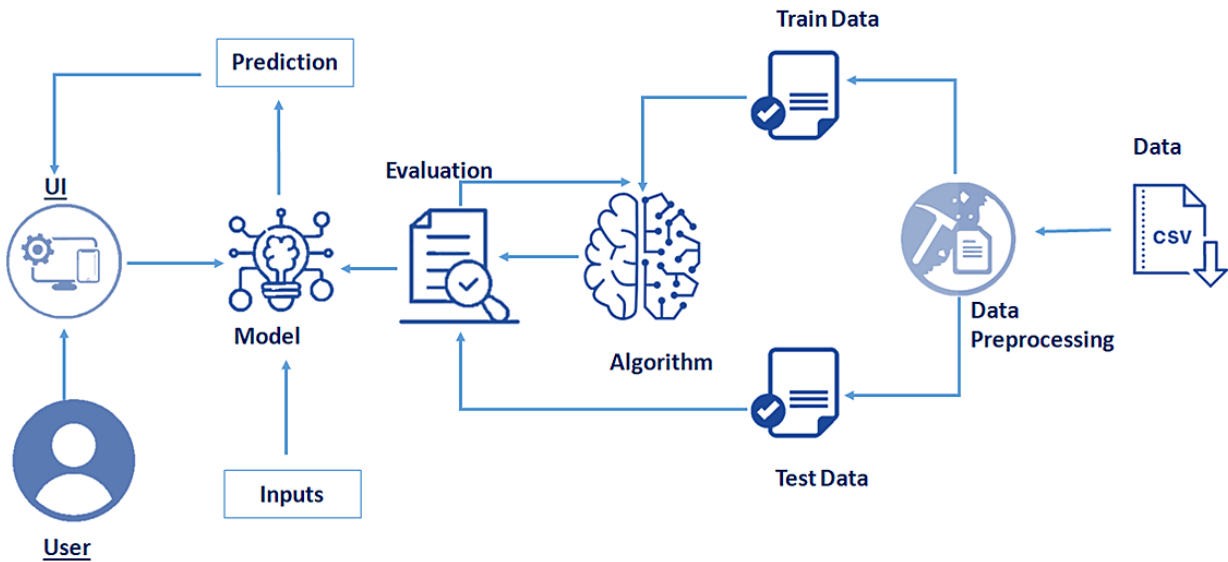


5.2 SOLUTION & TECHNICAL ARCHITECTURE

TECHNICAL ARCHITECTURE



SOLUTION ARCHITECTURE



5.3 USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Web user)	Login	USN-1	As a user, I can register for the application by entering my email.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I Will login,confirmation email once I have registered	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can login for the application through Mobile phone	i can login and access the application	Medium	Sprint-2
	Dashboard	USN-4	As a user, I need to enter my Details	I can get information as per details	High	Sprint-1
	Dashboard	USN-5	As a user, I need to enter my test details	I can get result based on test details	High	Sprint-1
Administrator	Services	USN-6	As a admin I need to provide valid result	I can get a result	High	Sprint-1
		USN-7	As an admin, I can add suggestions.	I can use it for later period	Medium	Sprint-3
Hospital Administrator	Login	USN-8	As an admin, I must collect input data for the medical database.	I can use for it further next step process	Medium	Sprint-3
	Dashboard	USN-9	As an admin, I need to login with appropriate access levels .	I can use for it further next step process	High	Sprint-1
Doctor/Radiologist	Diagnosis	USN-10	As a radiologist/doctor, I can view the diagnosis/ prediction results .	I can view the diagnosis/ prediction.	High	Sprint-2

6. PROJECT PLANNING & SCHEDULING

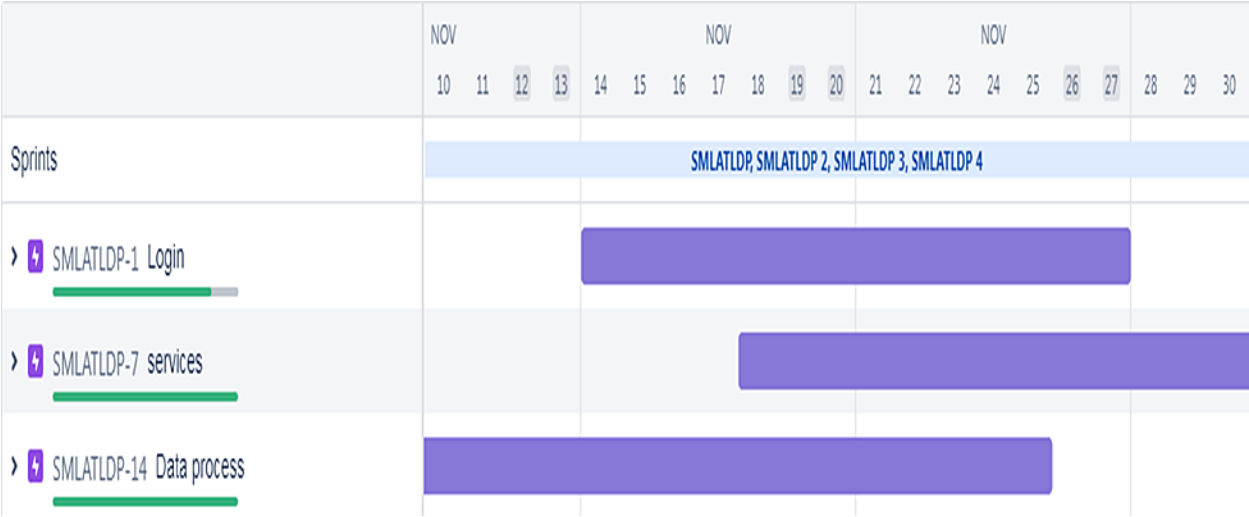
a. SPRINT PLANNING & ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Login	USN-1	As a user, I can login for the application by entering my mail	1	High	VINOTH J
Sprint-1		USN-2	As a user, I will login and get confirmation mail once I have registered	2	High	JENIFER A
Sprint-2		USN-3	As a user, I can login for the application through mobile number	2	Medium	AARTHI R
Sprint-2	services	USN-4	As a user, I need to enter my details	1	High	SESHAN M
Sprint-2		USN-5	As a user, I need to provide my Test Details	2	High	JAYAPRIYA R
Sprint-3	Data process	USN-6	As a admin I need to provide valid result	3	High	JAYAPRIYA R
Sprint-3		USN-7	As a admin I need to provide valid /useful suggestions	6	Medium	JENIFER A
Sprint-4	Login	USN-8	As a admin need to collect all the details and information.	2	High	AARTHI R
Sprint-4		USN-9	As a admin I need to store all the details and information	3	High	VINOTH J

b. SPRINT DELIVARABLE SCHEDULE

Sprint	Total Story Points	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



7. CODING & SOLUTIONING

a. FEATURE 1

Liver Patient Analysis

HomeGoto Predict

Introduction

Liver diseases averts the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises. Early prediction of liver disease using classification algorithms is an efficacious task that can help the doctors to diagnose the disease within a short duration of time. Discovering the existence of liver disease at an early stage is a complex task for the doctors. The main objective of this paper is to analyse the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease. This paper focuses on the related works of various authors on liver disease such that algorithms were implemented using Weka tool that is a machine learning software written in Java. Various attributes that are essential in the prediction of liver disease were examined and the dataset of liver patients were also evaluated. This paper compares various classification algorithms such as Random Forest, Logistic Regression and Separation Algorithm with an aim to identify the best technique. Based on this study, Random Forest with the highest accuracy outperformed the other algorithms and can be further utilised in the prediction of liver diseaserecommended

7.2 FEATURE 2

In the human body one of the most important organs is liver. If the regular functionality of the liver is disturbed then this condition is called disease affected liver. Therefore, an early stage of disease detection is more important which helps in disease prevention at starting stage with small medications. But, it is too difficult to identify Liver disease at early stages because symptoms are very less at the starting stage. Lab results with physical examination are involved in the Traditional methods. This paper aims to represent a Diagnosing for Liver disease prediction in Patients using Combined Machine Learning Models.

Liver Patient Prediction

Age:	Gender:
<input type="text"/>	<input type="text" value="Enter 0 as male, 1 as female"/>
Total_Bilirubin:	Direct_Bilirubin:
<input type="text"/>	<input type="text"/>
Alkaline_Phosphotase:	Alamine_Aminotransferase:
<input type="text"/>	<input type="text"/>
Aspartate_Aminotransferase:	Total_Protiens:
<input type="text"/>	<input type="text"/>
Albumin:	Albumin_and_Globulin_Ratio:
<input type="text"/>	<input type="text"/>
<input type="button" value="Predict"/>	

7.3 DATABASE SCHEMA

No uniform definition of Post his pate to my liver failure has been established in the literature addressing hepatic surgery. Considering the normal postoperative course of serum bilirubin concentration and International Normalized Ratio, we propose defining posthepatectomy liver failure as the impaired ability of the liver to maintain its synthetic, excretory, and detoxifying functions, which are characterized by an increased

international normalized ratio and concomitant hyperbilirubinemia (according to the normal limits of the local laboratory) on or after postoperative day 5. The severity of posthepatectomy liver failure should be graded based on its impact on clinical management. Grade a posthepatectomy liver failure requires no change of the patient's clinical management. The clinical management of patients with grade B posthepatectomy liver failure deviates from the regular course but does not require invasive therapy. The need for invasive treatment defines grade C posthepatectomy liver failure.

Liver Patient Prediction

You have a liver disease problem, You must and should consult a doctor. Take care

8. TESTING

8.1 TEST CASES

Steps To Execute	Test Data	Expected Result	Actual Result	Status	Comments	TC for Automation(Y/N)	BUG ID	Executed By
1.Enter the URL in web browser and click go 2.Verify home page displayed or not	http://127.0.0.1:5000	Home page should display	Working as expected	Pass		Yes		manual
1.Enter URL and click go 2.Verify home page display below UI elements: a.Home b.About c.Contact d.Get started	http://127.0.0.1:5000	Application should show below UI elements: a.Home b.About c.Contact d.Get started	Working as expected	pass		Yes		manual
1.Enter the URL (http://127.0.0.1:5000) 2.Click the About button 3.Verify About page displayed or not	http://127.0.0.1:5000	User should navigate to about page	Working as expected	pass		Yes		manual
1.Enter the URL (http://127.0.0.1:5000) 2.Click the "Get started" button 3.verify phishing website detection page displayed or not	click the get started button	User should navigate to phishing website detection page	Working as expected	Pass		Yes		manual

1.Enter the URL (http://127.0.0.1:5000) 2.Click the About button 3.Click the "Check your website" button in the About page 4.Verify phishing website detection page displayed or not	click the "check your website" button	user should navigate to phishing website detection page	Working as expected	Pass	Here user click the "check your website" button in about page	Yes		manual
1.Enter the URL (http://127.0.0.1:5000) 2.Click the About button 3.Click the "Check your website" button in the About page 4.enter the URL in the Phishing website detection page 5.click the predict button 6.verify it shows whether the URL entered by the user is safe or not	https://portal.naanmudhalvan.in.gov.in/login	Application should display "you are safe!! This is a legitimate website"	Working as expected	Pass	user enter the URL in correct format	Yes		Automatic
1.Enter the URL (http://127.0.0.1:5000) 2.Click the About button 3.Click the "Check your website" button in the About page 4.enter the URL in the Phishing website detection page 5.click the predict button 6.verify it shows whether the URL entered by the user is safe or not	https://www.searchonlinainfo.com/	Application should display "you are on the wrong site. Be cautious!"	Working as expected	Pass	User entered the URL in correct format	Yes		Automatic
1.Enter the URL (http://127.0.0.1:5000) 2.Click the About button 3.Click the "Check your website" button in the About page 4.enter the URL in the Phishing website detection page 5.click the predict button 6.verify it shows whether the URL entered by the user is safe or not	portal.naanmudhalvan.in.gov.in/login	Application should display "you are safe!! This is a legitimate website"	Not Working as expected	Fail	User enter the URL in correct format	Yes		Automatic

8.2 USER ACCEPTANCE TESTING

1. PURPOSE OF DOCUMENT

The purpose of this document is to briefly explain the test coverage and open issues of the project at the time of the release to User Acceptance Testing (UAT).

2. DEFECT ANALYSIS

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved.

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

3 .TEST CASE ANALYSIS

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

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	7	0	0	7
Client Application	50	2	0	48
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

9. RESULTS

9.1 PERFORMANCE METRICS

Finally, we need to check to see how well our model is performing on the test data.

Evaluation Metrics:

Accuracy score of SVM is

```
# Checking for accuracy score from actual data and predicted data
SVMaccuracy=accuracy_score(SVMpred, ytest)
SVMaccuracy

0.7606837606837606
```

Accuracy score of Random forest classification is

```
#Random Forest Classifier Model
from sklearn.ensemble import RandomForestClassifier
RFmodel=RandomForestClassifier()

# train the data with Random Forest model
RFmodel.fit(xtrain, ytrain)

RandomForestClassifier()

RFpred=RFmodel.predict(xtest)

# Checking for accuracy score from actual data and predicted data
RFaccuracy=accuracy_score(RFpred, ytest)
RFaccuracy

0.7094017094017094

# showing the confusion matrix
RFcm=confusion_matrix(RFpred, ytest)
RFcm

array([[77, 22],
       [12,  6]], dtype=int64)
```

Accuracy score of KNN classification is

```
# K-Nearest Neighbors Model
from sklearn.neighbors import KNeighborsClassifier
KNN = KNeighborsClassifier()

# train the data with K-Nearest Neighbors Model
KNN.fit(xtrain, ytrain)

KNeighborsClassifier()

KNNpred=KNN.predict(xtest)

# Checking for accuracy score from actual data and predicted data
KNNaccuracy=accuracy_score(KNNpred, ytest)
KNNaccuracy

0.6495726495726496

# showing the confusion matrix
KNNcm=confusion_matrix(KNNpred, ytest)
KNNcm

array([[70, 22],
       [19,  6]], dtype=int64)
```

As we can see that the accuracy score of the Support vector machine is higher compare to KNN and Random forest algorithms, we are proceeding with the support vector machine model.

1. ADVANTAGES

1. Those with liver disease may qualify for Social Security disability.
2. If you suffer from autoimmune hepatitis, cirrhosis, and other chronic liver conditions, you may qualify for disability benefits if the condition meets the Social Security blue book's listing.

11. DISADVANTAGES

1. Eating large amounts of liver can lead to symptoms of vitamin A toxicity.
2. Your own liver cannot process the excess vitamin A quickly enough, so eating a significant amount of liver regularly might lead to hyper vitamin is A.

CONCLUSION

1. The principal part of this work is to make an effective diagnosis
2. system for chorionic liver infection patients utilizing six distinctive
3. supervised machine learning classifiers. We researched all
4. classifiers execution on patient's information parameters and the
5. LR classifier gives the most elevated order exactness 75%
6. dependent on F1 measure to predict the liver disease and NB
7. gives the least precision 53%. From now on, the outperform
8. classification procedure will give for the decision support system
9. and diagnosis of chronic disease. The application will have the
10. option to predict liver infection prior and advise the wellbeing
11. condition. This application can be surprisingly gainful in low-salary
12. nations where our absence of medicinal foundations and just as
13. particular specialists. In our study, there are a few bearings for
14. future work in this field. We just explored some popular
15. supervised machine learning algorithms, more algorithms can be
16. picked to assemble an increasingly precise model of liver disease
17. prediction and performance can be progressively improved.
18. Additionally, this work likewise ready to assume a significant role
19. in health care research and just as restorative focuses to
20. anticipate liver infection

21. The principal part of this work is to make an effective diagnosis
22. system for chorionic liver infection patients utilizing six distinctive
23. supervised machine learning classifiers. We researched all
24. classifiers execution on patient's information parameters and the
25. LR classifier gives the most elevated order exactness 75%
26. dependent on F1 measure to predict the liver disease and NB
27. gives the least precision 53%. From now on, the outperform
28. classification procedure will give for the decision support system
29. and diagnosis of chronic disease. The application will have the
30. option to predict liver infection prior and advise the wellbeing
31. condition. This application can be surprisingly gainful in low-salary
32. nations where our absence of medicinal foundations and just as
33. particular specialists. In our study, there are a few bearings for
34. future work in this field. We just explored some popular
35. supervised machine learning algorithms, more algorithms can be
36. picked to assemble an increasingly precise model of liver disease
37. prediction and performance can be progressively improved.
38. Additionally, this work likewise ready to assume a significant role
39. in health care research and just as restorative focuses to
40. anticipate liver infection

The principal part of this work is to make an effective diagnosis system for chorionic liver infection patients utilizing six distinctive supervised machine learning classifiers. We researched all classifiers execution on patient's information parameters and the LR classifier gives the most elevated order exactness 75% dependent on F1 measure to predict the liver disease and NB gives the least precision 53%. From now on, the outperform classification procedure will give for the decision support system and diagnosis of chronic disease. The application will have the option to predict liver infection prior and advise the wellbeing condition. This application can be surprisingly gainful in low-salary nations where our absence of medicinal foundations and just as particular specialists. In our study, there are a few bearings for future work in this field. We just explored some popular supervised machine learning algorithms, more algorithms can be picked to assemble an increasingly precise model of liver disease prediction and performance can be progressively improved. Additionally, this work likewise ready to assume a significant role in health care research and just as restorative focuses to anticipate liver infection

The principal part of this work is to make an effective diagnosis system for chorionic liver infection patients utilizing six distinctive supervised machine learning classifiers. We researched all

classifiers execution on patient's information parameters and the LR classifier gives the most elevated order exactness 75% dependent on F1 measure to predict the liver disease and NB gives the least precision 53%. From now on, the outperform classification procedure will give for the decision support system and diagnosis of chronic disease. The application will have the option to predict liver infection prior and advise the wellbeing condition. This application can be surprisingly gainful in low-salary nations where our absence of medicinal foundations and just as particular specialists. In our study, there are a few bearings for future work in this field. We just explored some popular supervised machine learning algorithms, more algorithms can be picked to assemble an increasingly precise model of liver disease prediction and performance can be progressively improved. Additionally, this work likewise ready to assume a significant role in health care research and just as restorative focuses to anticipate liver infection

11. FEATURE SCOPE

With a growing trend of sedentary and lack of physical activities, diseases related to liver have become a common encounter nowadays. In rural areas the intensity is still manageable, but in urban areas, and especially metropolitan areas the liver disease is a very common sighting nowadays. Liver diseases cause millions of deaths every year. Viral hepatitis alone causes 1.34 million deaths every year. Problems with liver patients are not easily discovered in an early stage as it will be functioning normally even when it is partially damaged. An early diagnosis of liver problems will increase Patients survival rate. Liver failures are at high rate of risk among Indians. It is expected that by 2025 India may become the World Capital for Liver Diseases. The widespread occurrence of liver infection in India is contributed due to deskbound lifestyle, increased alcohol consumption and smoking. There are about 100 types of liver infections.

12. APPENDIX

Source Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

[ ] data = pd.read_csv("/content/archive.zip")

[ ] data.head()

[ ] data.tail()

data.info()

[ ] data.describe()

[ ] data.isnull().sum()

[ ] data['Albumin_and_Globulin_Ratio']=data['Albumin_and_Globulin_Ratio'].fillna(data['Albumin_and_Globulin_Ratio'].median())
data.isnull().sum()

[ ] sns.countplot(data =data,x ='Gender',label ='Count')

[ ] sns.countplot(data = data, x = 'Dataset')

[ ] x=data.iloc[:,0:-1]
y=data.iloc[:, -1]

[ ] from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2)

[ ] xtrain.shape

[ ] xtest.shape
```



```
[ ] from sklearn.ensemble import RandomForestClassifier
    RFmodel=RandomForestClassifier()
```

```
[ ] RFmodel.fit=(xtrain, ytrain)
```

```
[ ] from sklearn.metrics import accuracy_score
    SVMaccuracy= accuracy_score(svm,ytest)
    SVMaccuracy
```

```
[ ] RFpred=RFmodel.predict=(xtest)
    RFaccuracy=accuracy_score(RFpred,ytest)
    RFaccuracy
```

```
[ ] RFcm=confusion_matrix=(RFpred,ytest)
    RFcm
```

```
[ ] from sklearn.neighbors import KNeighborsClassifier
    KNN=KNeighborsClassifier()
```



```
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
```

```
[ ] svm=SVC()
    RFmodel=RandomForestClassifier()
    KNNmodel= KNeighborsClassifier()
```

```
[ ] from sklearn.svm import SVC
    svm=SVC()
```

```
[ ] xtrain
```

```
[ ] ytrain
```

```
[ ] svm.fit(xtrain,ytrain)
```

```
[ ] KNN.fit=(xtrain,ytrain)
```

```
[ ] KNNpred=KNN.predict=(xtest)
```

```
[ ] KNNaccuracy=accuracy_score(KNNpred,ytest)
KNNaccuracy
```

```
▶ KNNcm=confusion_matrix=(KNNpred,ytest)
KNNcm
```

```
[ ] import pickle
pickle.dump(svm,open('liver_analysis.pkl','wb'))
```

Home.html

<Html>

<Body>

<h1 style="background-color: violet ;"> Liver Patient Analysis</h1>

<h2 style="background-color: yellow ;">

<p> Liver diseases avert the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises. Early prediction of liver disease using classification algorithms is an efficacious task that can help the doctors to diagnose the disease within a short duration of time. Discovering the existence of liver disease at an early stage is a complex task for the doctors. The main objective of this project is to analyze the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease. This paper focus on the related works of various authors on liver disease such that algorithms where implemented using Wake tool that is a machine learning software written in java. Various attributes that are essential in the prediction of liver disease where examined and the dataset of liver patients were also evaluated. This paper compares various classification algorithm such as Random Forest, Logistic Regression and Separation Algorithm with an aim to identify the best technique based on Study, Random Forest with the highest accuracy outperformed the other algorithms and can be further utilized in the prediction of liver disease recommended.

</p>

</html>

</body>

Index.html

```
<html>
<body>
<h1 style="background-color: violet;">Liver Patient prediction</h1>
  <h2 style="background-color: yellow;">
    <td>
      <tr>
<table width="50"><td>
  <label for="name">Age:</label><br><br>
  <input type="text" id="name" name="name"><br>
</td>
  <td width="50">
    <td>
      <label for="name">Gender:</label><br><br>
      <input type="text" id="name" name="name">
    </td>
  </tr>
</tr>
<table width="50"><td>
  <label for="name">Total_Bilirubin:</label><br> <br>
  <input type="text" id="name" name="name"><br>
</td>
  <td width="50">
    <label for="name">Direct_Bilirubin:</label><br><br>
    <input type="text" id="name" name="name">
```

```

<input type="text" id="name" name="name">
</td>
</tr>

<tr>
<table width="50"><td>
  <label for="name">Alkaline_Phosphatase:</label><br> <br>
  <input type="text" id="name" name="name"><br>
</td>
<td width="50">
  <label for="name">Alamine_Aminotransferase:</label><br><br>
  <input type="text" id="name" name="name">
</td>
</tr>

<tr>
<table width="50"><td>
  <label for="name">Aspartate_Aminotransferase:</label><br><br>
  <input type="text" id="name" name="name"><br>
</td>
<td width="50">
  <label for="name">Total_Protiens:</label><br><br>
  <input type="text" id="name" name="name">

  </td>
</tr>
<td>
  <table width="50"><td>
    <label for="name">Albumin:</label><br><br>
    <input type="text" id="name" name="name"><br><br>
    <form action="predict">
      <input type="submit" value="predict" />
    </form>

  <td width="50">
    <label for="name">Albumin_and_Globulin_Ratio:</label><br><br>
    <input type="text" id="name" name="name">
  </td>
</tr>

</body>
</html>

```

DEMO LINK

<https://github.com/IBM-EPBL/IBM-Project-40052-1660622281/blob/main/Final%20Delivarables/Project%20Demo.mp4>