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 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 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.

1.2 PURPOSE

This project helps in early risk identification and thus increasing the chance of recovery of the patient as medical attention is given early on. No prior knowledge of Medical Sciences and Liver Diseases is needed as we only need the medical report and we need to enter them in the GUI. The results are instantaneous thus reducing the time of diagnosis. The system predicts the result with high accuracy thus it is trustworthy at a large scale. This Liver Prediction System could be used in hospitals, so before sending the samples to the lab the hospital could eliminate False Negative Prediction and could start the medication course early on, thus saving valuable time. As this system is integrated into a flask-based web application many people can use this and thus increasing the impact of our project.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

The Machine Learning Model is yet to be integrated with a Web/Mobile Application, thus limiting the number of users and reducing the accessibility. The Final result is binary ('Risk' or 'No Risk') thus limiting the scope of diagnosis. There isn't a disease-specific model that will help in early predictions of different kinds of liver diseases.

2.2 REFERENCES

S.No	AUTHOR	TITLE	YEAR
1	Chokka Anuradha, D Swapna, Balamuralikrishnan	Diagnosing of Liver Disease Prediction in Patients using combined Machine Learning Models	IEEE-2022
2	Fazle Rabbi, S.M. Mahedy Hasan, Arifa Islam Champa, Kamrul Hasan	Prediction of Liver Disease Using Machine Learning Algorithms: A Comparative Study	ICAICT-2020
3	Varun Vats, Lining Zhang	A Comparative Analysis of Unsupervised Machine Techniques for Liver Disease Prediction	IEEE-2018
4	A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim	A Comparative Study on Liver Disease Prediction Using Supervised Machine Learning Algorithms	ResearchGate-2019
5	Rakshith D B, Mrigank Srivastava, Ashwani Kumar, Gururaj S P	Liver Disease Predicition System Using Machine Learning Techniques	IJERT-2021

2.3 Problem Statement Definition

1.Diagnosing of Liver Disease Prediction in Patients using combined Machine Learning Models

Chokka Anuradha, D Swapna, Balamuralikrishnan Thati

IEEE-2022

In the human body one of the most important organs is the 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 the 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. Optimized three machine learning algorithms are used for the accurate diagnosis of liver disease and they are Artificial Neural Networks (ANN), Decision Trees, and K-Nearest Neighbors (KNN). With the help of these algorithms, given data is classified and results are produced. The future data is predicted with the help of past and present data. The accuracy results are produced by comparing three classification algorithms

2.Prediction of Liver Disease Using Machine Learning Algorithms: A Comparative Case Study

Fazle Rabbi, S. M. Mahedy Hasan, Arifa Islam Champa, Kamrul Hasan.

ICAICT-2020

Liver, a crucial interior organ of the human body whose principal tasks are to eliminate generated waste produced by our organism, digest food, and preserve vitamins and energy materials. The liver disorder can cause various fatal diseases, including liver cancer. Here they have compared four different ML algorithms such as Logistic Regression, Decision Tree, Random Forest and Extra Trees for classifying Indian Liver Patient Dataset. Pearson Correlation

Coefficient based features selection is applied to eliminate irrelevant features from the dataset. After comparing experimental results, we have found that boosting on ET provides the highest accuracy of 92.91%

3. A Comparative Analysis of Unsupervised Machine Techniques for Liver Disease Prediction

Varun Vats, Lining Zhang

IEEE-2018

He considered three different ML (Machine Learning) algorithms. A comparison of these algorithms had been carried out for evaluating their forecasting accuracy and computing intricacy \. These algorithms included AP (Affinity Propagation), K means and DBSCAN. This work was dedicated to the medical dataset based on lever disorders. This work made use of the Silhouette coefficient to measure the comparative efficiency of the considered algorithmic approaches

4. A Comparative Study on Liver Disease Predicition Using Machine Learning Algorithms

A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain

ResearchGate – 2019

Chronic Liver Disease is the leading cause of global death that impacts the massive quantity of humans around the world. This disease diagnosis is very costly and complicated. Therefore, this paper evaluates the performance of different Machine Learning algorithms in order to reduce the high cost of chronic liver disease diagnosis by prediction. Six machine learning techniques have been applied including Logistic Regression, K Nearest Neighbors, Decision Tree, Support Vector Machine, Naïve Bayes, and Random Forest. The performance was evaluated on different measurement techniques such as accuracy, precision, recall, f-1 score, and specificity and the result was that LR achieved the highest accuracy.

5. Liver Disease Prediction System using Machine Learning Techniques

Rakshith D B, Mrigank Srivastava, Ashwani Kumar, Gururaj S P

IJERT – 2021

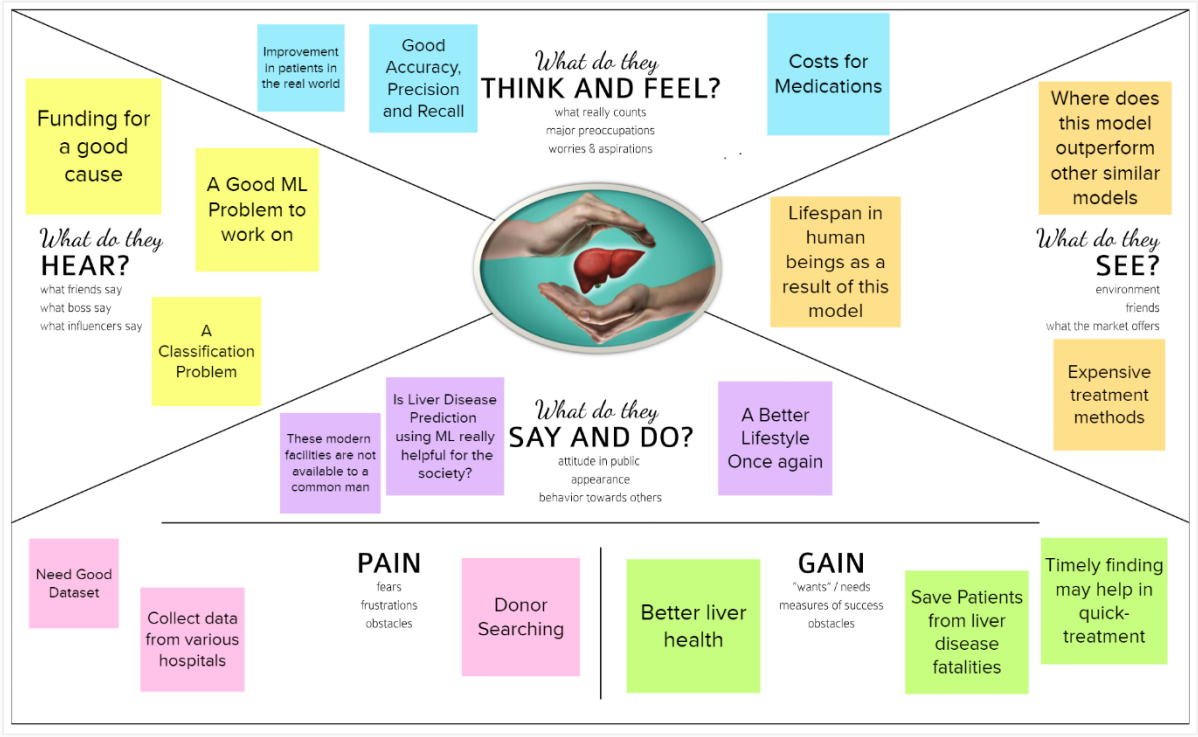
In this paper risk of liver disease for a person is predicted based on the blood test report results of the user. With the dataset used for this project, 100 % accuracy is obtained for SVM model. The data pre-processing was done using Jupyter Notebook and Desktop Application was Implemented using Spyder IDE. The programming language which was used is python and machine learning Sklearn was used to build the model using classification algorithm like KNN, SVM, Naive Bayes and ANN.

3. IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

Statistical Machine Learning Approaches To Liver Disease Prediction

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.



3.2 IDEATION AND BRAINSTORMING

TOP 3 POINTS:

1. Through this project, we can successfully develop and validate a risk prediction model and subsequent user-friendly scoring tool, the Algorithm for Liver Function Investigations, for liver condition diagnosis in patients with no obvious liver condition at the time of incident liver function testing in primary care.
2. Upon early diagnosis of the liver condition, the severity would reduce and so does the cost of appropriate treatment methodologies. Hence, this project will help reduce the financial as well as biological impact.
3. Liver cirrhosis is a leading cause of death and affects millions of people in the United States. Early mortality prediction among patients with cirrhosis might give healthcare providers more opportunity to effectively treat the condition.

DISCUSSION BETWEEN GROUP MEMBERS:

Balakumaran S:

1. Early risk identification.
2. Saving cost and time.
3. Liver is an important organ in our human body where most of the metabolism takes place.
4. Can help many people as its mortality rate is very high and affects millions of people.

Saravanan P:

5. Additional enzymes or proteins that affect the liver can be included.

6. Based on the analytics we can analyse which patients are most likely to suffer from heart disease in the near future and based on the patient details we will make decisions to cure them.

7. To detect disease, Health Care Professionals need to collect samples from patients which can cost both time and money. The main problem is doctors cannot diagnose on the basis of variations in test results.

8. Current method of doctors analysing each blood report is difficult. ML models can help speed up the process.

Buvaneswar AS:

9. Analyse the correlation between the different features of the initial-model.

10. What is the relation between the severity of the liver disease w.r.t age.

11. Can we somehow collect feedback from patients to identify any underlying relation b/w potential lifestyles that impact the liver in a very substantial way.

12. Talk with doctors to gather ideas about different data sources available for building the model.

Abhinav K:

13. Identification of different features that may help in the building of the model.

14. Analysing what tests are taken when a patient undergoes diagnosis of liver and accordingly use those set of details in our model as applicable.

15. Develop disease-specific models that will help in early predictions of different kinds of liver diseases.

16. Perform an EDA to see if there are any common features between the disease-specific models.

3.3 PROPOSED SOLUTION:

S.No	Parameter	Description
1	Problem Statement (Problem to be solved)	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 problem at hand 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.
2	Idea / Solution description	Through this project, we can successfully develop and validate a risk prediction model and subsequent user-friendly scoring tool, the Algorithm for Liver Function Investigations, for liver condition diagnosis in patients with no obvious liver condition at the time of incident liver function testing in primary care. Upon early diagnosis of the liver condition, the severity would reduce and so does the cost of appropriate treatment methodologies. Hence, this project will help reduce the financial as well as biological impact. Liver cirrhosis is a leading cause of death and affects millions of people in the United States. Early mortality prediction among patients with cirrhosis might give healthcare providers more opportunity to effectively treat the condition.
3	Novelty / Uniqueness	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. We have tried to minimize the number of False Negative predictions.
4	Social Impact / Customer Satisfaction	This project helps in early risk identification and thus increasing the chance of recovery of the patient as medical attention is given early on. No prior knowledge of Medical Sciences and Liver Diseases is needed as we only need the medical report and we need to enter them in the GUI. The results are instantaneous thus reducing the time of diagnosis. The system predicts the result with high accuracy thus it is trustworthy at a large scale.
5	Business Model (Revenue Model)	This could be made as a Proprietary Software and could be sold to hospital chains i.e., integrate this Liver Disease Prediction system to their main frame. So before sending the samples to the lab the hospital could eliminate False Negative Prediction and could start the medication course before the test results arrive saving valuable time. This could also be made available to the general public and if a patient has the symptoms of a liver disease, they could get a clear idea of their condition by using the web app, thus giving us valuable market penetration.
6	Scalability of the Solution	The Liver Disease Prediction System is integrated into a flask-based web application. Thus, many people could use it simultaneously thus increasing the impact of our project. As there is a user-friendly GUI, the project will be accessible to a large audience regardless of their educational background.

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	Registration through Form
FR-2	User Verification	Submit necessary medical documents for validation
FR-3	User Dashboard	Redirect to dashboard after successful login and put a hyperlink to enter the data.
FR-4	Follow-up	Contact the user to see if they got in touch with the recommended doctors around their region. (Assuming they have been predicted to have the disease)
FR-5	Database	Create a database to story all activity that takes place in the web app for administrative/development purposes.

Following are the functional requirements of the proposed solution.

4.2 NON-FUNCTIONAL REQUIREMENTS:

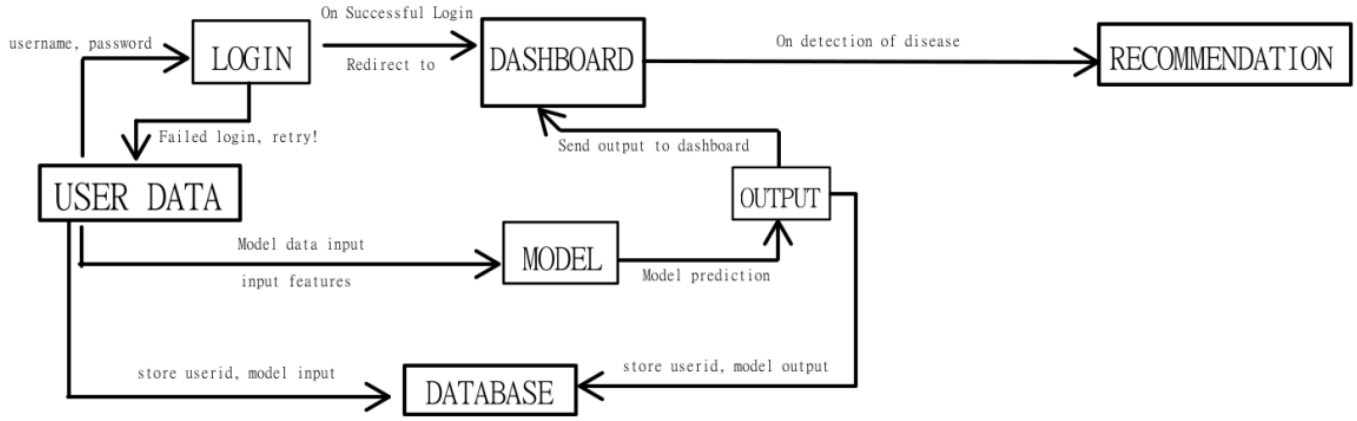
Following are the non-functional requirements of the proposed solution.

FR No.	NON-FUNCTIONAL REQUIREMENTS	DESCRIPTION
NFR-1	Usability	A functioning server that is capable of running the application 24x7
NFR-2	Security	Follow valid general-purpose protocols for providing security and privacy to the data and users.
NFR-3	Reliability	Tweak the model as and when new findings arise to make the model as reliable as possible.
NFR-4	Performance	Make the whole process as user-friendly as possible
NFR-5	Availability	Model is preferred to be available 24x7 as the application to be hosted on a web server.
NFR-6	Scalability	Model can be tweaked and easily changed when new findings can prove useful insights for getting a better model. Can also change the design of the website as needed.

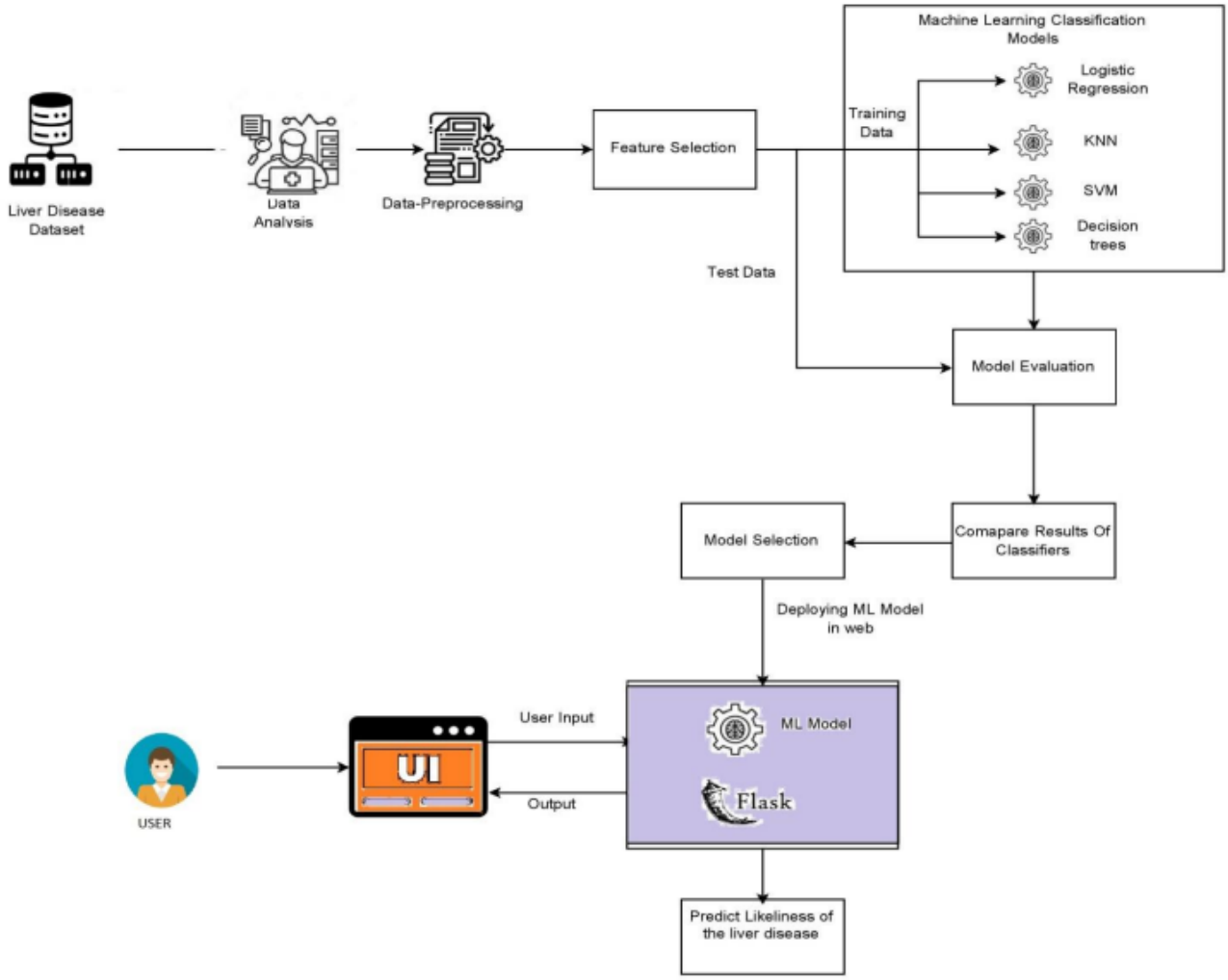
5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 SOLUTION ARCHITECTURE



5.3 USER STORIES

USER TYPE	FUNCTIONAL REQUIREMENT	USER STORY NO.	USER STORY	ACCEPTANCE CRITERIA	PRIORITY	RELEASE
Customer (web user)	Registration	USN-1	As a user, I can register for the application by entering my username, State, password, and confirming my password	I can access my account / dashboard	High	Sprint-1

	Login	USN-2	As a user, I can log into the application by entering username & password	I can access the dashboard of the application	High	Sprint-2
	Dashboard	USN-3	As a user, I can view the dashboard and find the page to enter my data for prediction	I can find the link to enter the data for prediction	High	Sprint-3
Customer with a risk of Disease	Follow-up	USN-4	As a Customer with risk of disease I was recommended 3 Liver Specialists from my State	I can find the Liver Specialists from my State	Low	Sprint-4

6.PROJECT PLANNING AND SCHEDULING

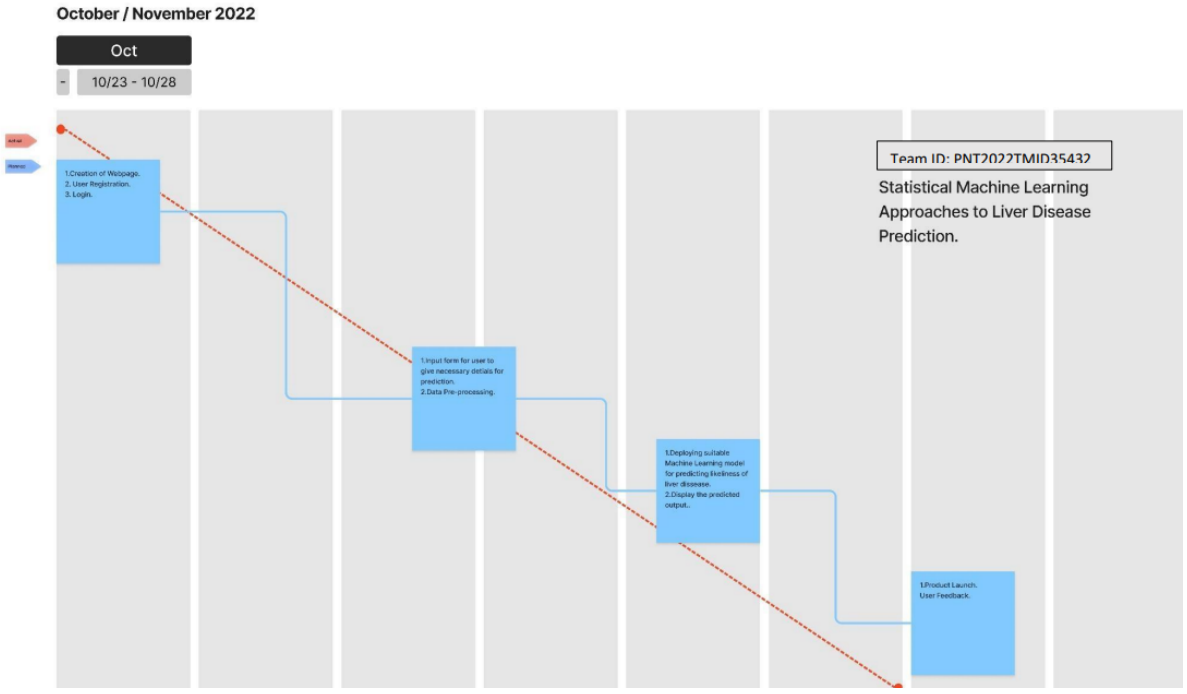
6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Functional Requirement	User Story No.	User Story	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	5	High	Abhinav K
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	5	High	Abhinav K
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email & password	10	High	Abhinav K
Sprint-2	Input Necessary Details	USN-4	As a user, I can give Input Details to Predict Likeliness of Liver Disease	15	High	Saravanan P
Sprint-2	Data Pre-Processing	USN-5	Transform raw data into suitable format for prediction.	5	High	Saravanan P
Sprint-3	Prediction of Liver Disease	USN-6	As a user, I can predict Liver Disease using machine learning model. (Setting up the web application with the appropriate *.pkl file)	15	High	Balakumaran S
Sprint-3		USN-7	As a user, I can get accurate prediction of liver disease. (Trying to test accuracy with different models)	5	Medium	Buvaneswar A S
Sprint-4	Review	USN-8	As a user, I can give feedback of the application.	20	High	Buvaneswar A S

6.2 SPRINT DELIVERY 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		
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022		

6.3 REPORTS FROM JIRA



7.CODING AND SOLUTIONING

7.1 Authentication

REGISTRATION:

```
class RegisterForm(FlaskForm):
    username = StringField(validators=[
                           InputRequired(), Length(min=4, max=20)], render_kw={"placeholder": "Username"})

    password = PasswordField(validators=[
                             InputRequired(), Length(min=2, max=20)], render_kw={"placeholder": "Password"})

    state = SelectField(
        label="Choose an option", choices=["Choose an option",'TAMIL NADU','KERLA','ANDRA PRADESH','TELANGANA','KARNATAKA',"MAHARASHTRA",'ODISHA','WEST BENGAL','RAJASTHAN','PUNJAB'] )

    submit = SubmitField('Register')

    def validate_state(self, state):
        state = str(state)
        if state[state.find("selected")+16:state.find("selected")+32] == "Choose an option":
            raise ValidationError("Please select a role.")

    def validate_username(self, username):
        existing_user_username = User.query.filter_by(
            username=username.data).first()
        if existing_user_username:
            raise ValidationError(
                'That username already exists. Please choose a different one.')
```

The Registration Page includes three forms namely Username, Password and State. It raises an error if the entered state is not valid or if the entered username already exists.

```
@ app.route('/register', methods=['GET', 'POST'])
def register():
    form = RegisterForm()
    if form.validate_on_submit():
        hashed_password = bcrypt.generate_password_hash(form.password.data)
        new_user = User(username=form.username.data,
                        password=hashed_password, state=form.state.data)
        db.session.add(new_user)
        db.session.commit()
        return redirect(url_for('login'))
    else:
        Warning('Login Unsuccessful. Please check username and password', 'danger')

    return render_template('register.html', form=form)
```

The details of the user are stored in a database. There are 3 columns in the database, the username (primary key), hashed password and state. After clicking the submit button in the registration page the user is redirected to the login page if there is no error raised.

LOGIN:

```
class LoginForm(FlaskForm):
    username = StringField(validators=[
                           InputRequired(), Length(min=4, max=20)], render_kw={"placeholder": "Username"})

    password = PasswordField(validators=[
                             InputRequired(), Length(min=8, max=20)], render_kw={"placeholder": "Password"})

    submit = SubmitField('Login')
```

The Login page has two forms, username and password.

```
@app.route('/', methods=['GET', 'POST'])
def login():
    form = LoginForm()
    if form.validate_on_submit():
        user = User.query.filter_by(username=form.username.data).first()
        if user:
            if bcrypt.check_password_hash(user.password, form.password.data):
                login_user(user)
                session['admin'] = request.form['username']
                session['state'] = user.query.filter_by(username=form.username.data).first().state
                return redirect(url_for('home'))

    return render_template('login.html', form=form)
```

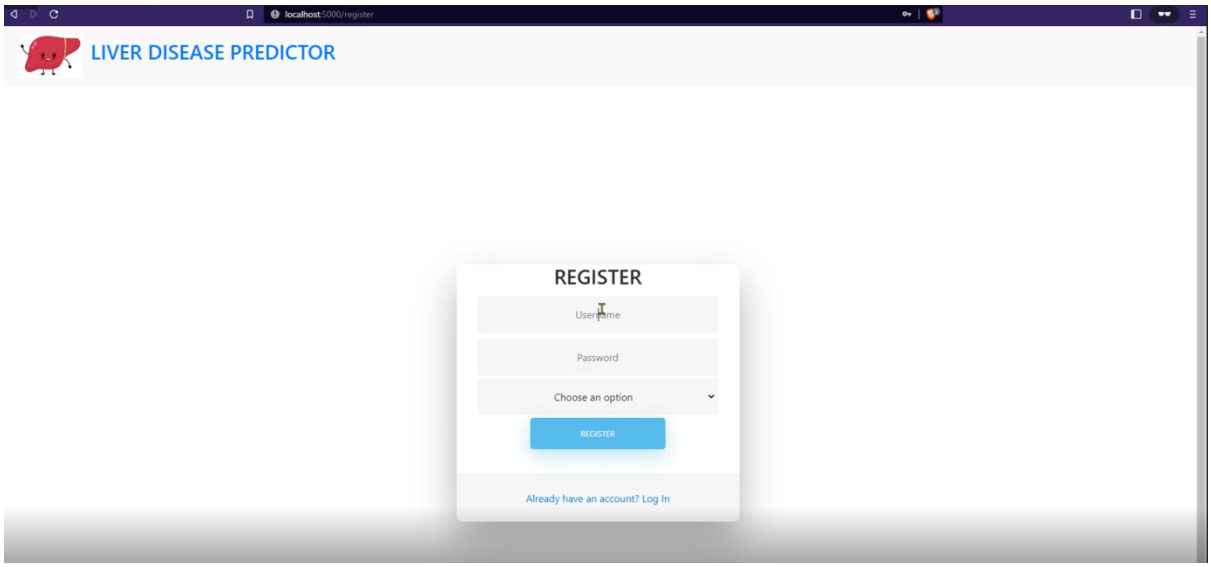
If the username is in the database and if the entered password is correct then the session admin is set to the username, the session state is the state of the logged in user and the user gets redirected to the home page.

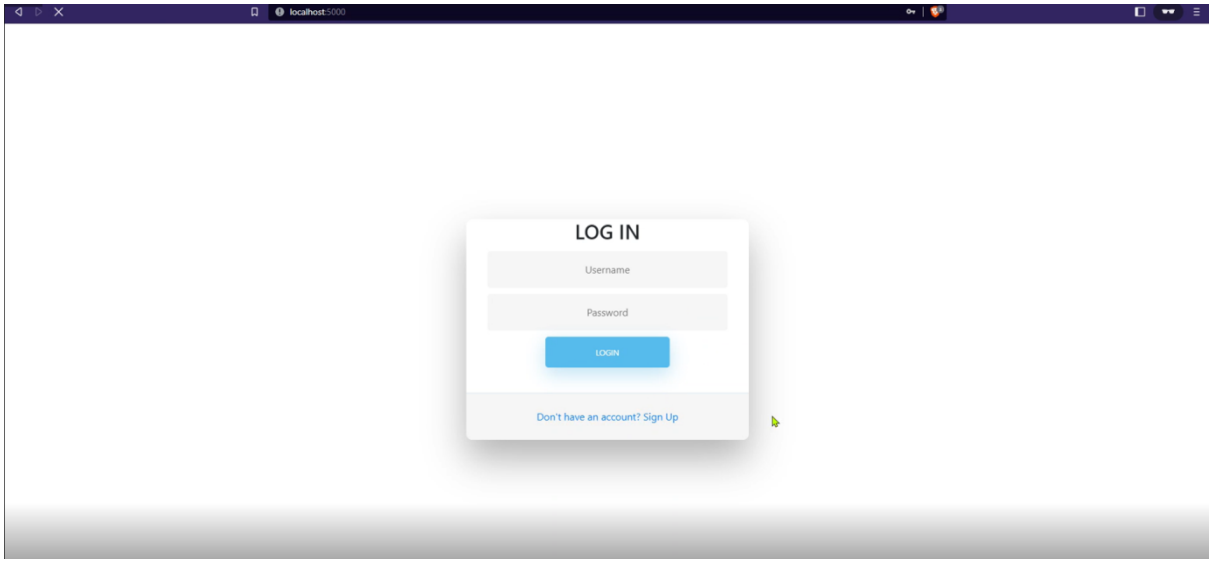
7.2 DASHBOARD

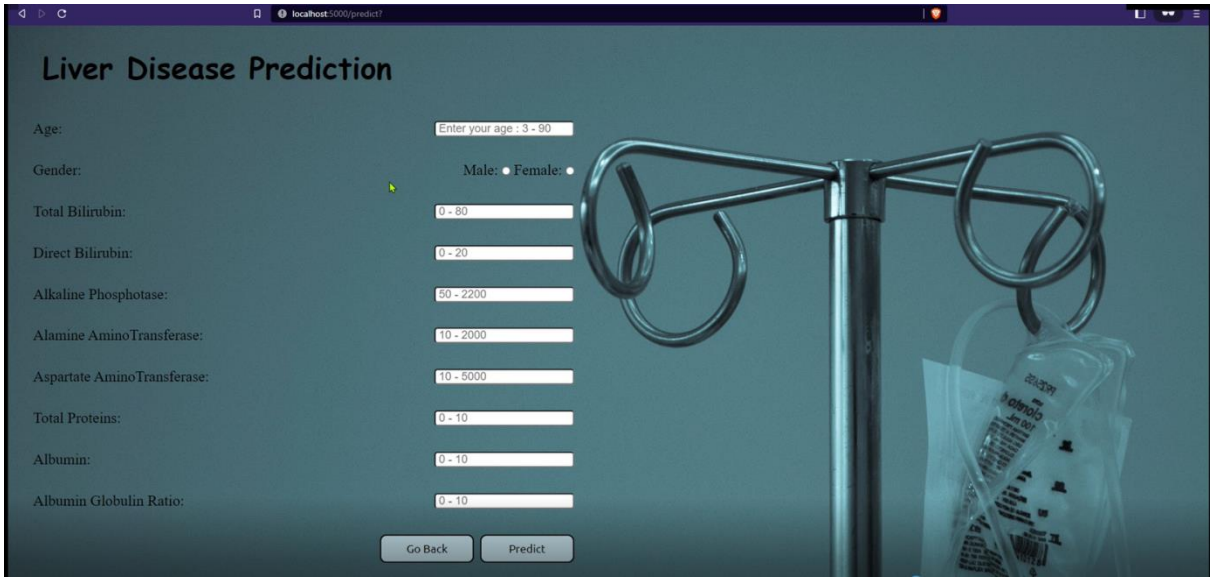
The data required by the model is entered by the user in the dashboard. Then the data is passed into the model. The results of the model are displayed and if the user has risk of Liver Disease, then three Hepatologists from their state are recommended for further treatment.

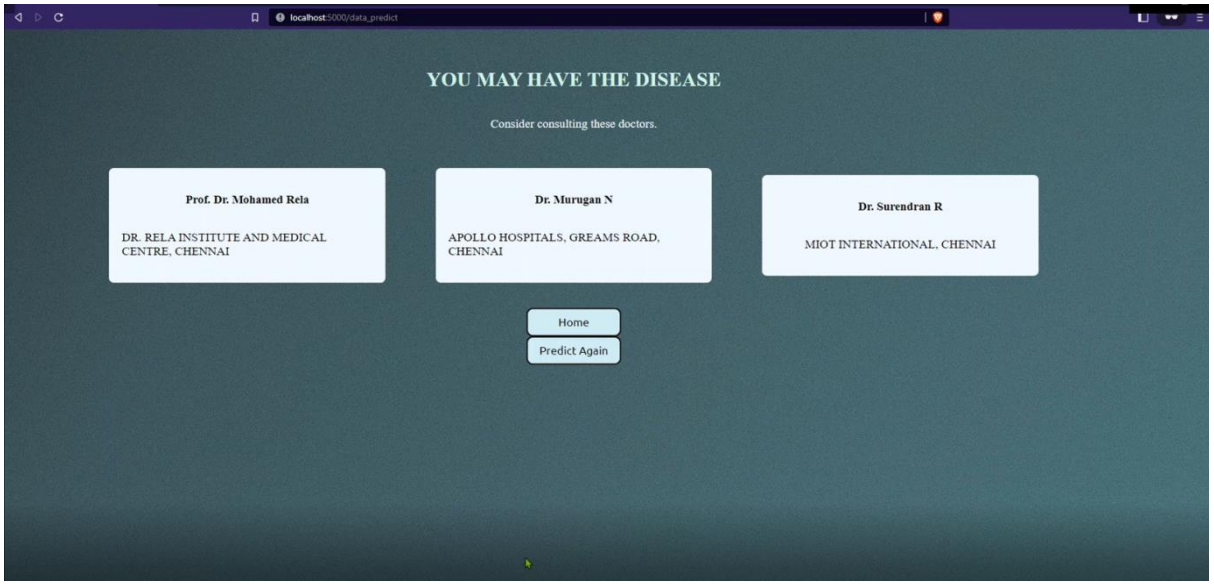
```python
@app.route('/data_predict', methods=['POST'])
def predict():

    age = request.form['age']
    gender = request.form['gender']
    tb = request.form['tb']
    dbi = request.form['dbi']
    ap = request.form['ap']
    aa1 = request.form['aa1']
    aa2 = request.form['aa2']
    tp = request.form['tp']
    a = request.form['a']
    agr = request.form['agr']
    if gender == "Male":
        gender = 1
    else:
        gender = 0
    data = [[float(age),
            float(gender),
            float(tb),
            float(dbi),
            float(ap),
            float(aa1),
            float(aa2),
            float(tp),
            float(a),
            float(agr)]]

    model = pickle.load(open('KNN_10thaug.pkl', 'rb'))

    prediction = model.predict(data)
    if (prediction == 1):
        return render_template('noChance.html')
    else:
        return render_template('chance.html', specialist = specialists.query.filter_by(state = session['state']).first())
```

OUTPUTS:




localhost:5000
LOG IN
Username
Password
LOGIN
Don't have an account? Sign Up


localhost:5000/predict?
Liver Disease Prediction
Age:
Enter your age : 3 - 90
Gender:
Male:
Female:
Total Bilirubin:
0 - 80
Direct Bilirubin:
0 - 20
Alkaline Phosphotase:
50 - 2200
Alamine AminoTransferase:
10 - 2000
Aspartate AminoTransferase:
10 - 5000
Total Proteins:
0 - 10
Albumin:
0 - 10
Albumin Globulin Ratio:
0 - 10
Go Back
Predict


localhost:5000/data_predict
YOU MAY HAVE THE DISEASE
Consider consulting these doctors.
Prof. Dr. Mohamed Rela
DR. RELA INSTITUTE AND MEDICAL CENTRE, CHENNAI
Dr. Murugan N
APOLLO HOSPITALS, GREAMS ROAD, CHENNAI
Dr. Surendran R
MIOT INTERNATIONAL, CHENNAI
Home
Predict Again

8.TESTING

8.1 USER ACCEPTANCE TESTING

TEST No.	FEATURE	TEST SCENARIO	EXPECTED RESULT	ACTUAL RESULT	STATUS
1	Main Page	Enter URL and click go.	Login Page should display	Working as Expected	Pass
2	Registration	1.Click Register button 2.Enter Registration details	1.Register Page should be displayed 2.User should be redirected to the login page after registering	Working as Expected	Pass
3	Login	Enter Login Details	User should be redirected to home page if login is successful	Working as expected	Pass
4	Home	Click Predict	1.The contents of the home page should be displayed. 2.User should be redirected to predict page	Working as expected	Pass
5	Predict	Enter User health record	1.Ensure the values are within in the permissible range. 2. The results of the model should be displayed after clicking predict	Working as expected	Pass
6	Results		Results of the model are displayed	Working as expected	Pass

9.RESULTS

ID	Name	Reference	Turbo
lr	Logistic Regression	sklearn.linear_model._logistic.LogisticRegression	True
knn	K Neighbors Classifier	sklearn.neighbors._classification.KNeighborsCl...	True
nb	Naive Bayes	sklearn.naive_bayes.GaussianNB	True
dt	Decision Tree Classifier	sklearn.tree._classes.DecisionTreeClassifier	True
svm	SVM - Linear Kernel	sklearn.linear_model._stochastic_gradient.SGDC...	True
rbfsvm	SVM - Radial Kernel	sklearn.svm._classes.SVC	False
gpc	Gaussian Process Classifier	sklearn.gaussian_process._gpc.GaussianProcessC...	False
mlp	MLP Classifier	sklearn.neural_network._multilayer_perceptron....	False
ridge	Ridge Classifier	sklearn.linear_model._ridge.RidgeClassifier	True
rf	Random Forest Classifier	sklearn.ensemble._forest.RandomForestClassifier	True
qda	Quadratic Discriminant Analysis	sklearn.discriminant_analysis.QuadraticDiscrim...	True
ada	Ada Boost Classifier	sklearn.ensemble._weight_boosting.AdaBoostClas...	True
gbc	Gradient Boosting Classifier	sklearn.ensemble._gb.GradientBoostingClassifier	True
lda	Linear Discriminant Analysis	sklearn.discriminant_analysis.LinearDiscrimina...	True
et	Extra Trees Classifier	sklearn.ensemble._forest.ExtraTreesClassifier	True
lightgbm	Light Gradient Boosting Machine	lightgbm.sklearn.LGBMClassifier	True
dummy	Dummy Classifier	sklearn.dummy.DummyClassifier	True

List of models trained for the given dataset.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lda	Linear Discriminant Analysis	0.7157	0.7125	0.9552	0.7291	0.8265	0.1053	0.1622	0.025
ridge	Ridge Classifier	0.7133	0.0000	0.9828	0.7180	0.8296	0.0453	0.0880	0.014
svm	SVM - Linear Kernel	0.7129	0.0000	0.8345	0.7911	0.8003	0.2523	0.2770	0.018
dummy	Dummy Classifier	0.7109	0.5000	1.0000	0.7109	0.8310	0.0000	0.0000	0.012
lr	Logistic Regression	0.7085	0.7353	0.9138	0.7379	0.8157	0.1434	0.1837	0.026
lightgbm	Light Gradient Boosting Machine	0.6960	0.6847	0.8345	0.7620	0.7955	0.2022	0.2085	0.045
et	Extra Trees Classifier	0.6888	0.7081	0.8621	0.7428	0.7972	0.1362	0.1431	0.462
rf	Random Forest Classifier	0.6813	0.7008	0.8276	0.7510	0.7867	0.1578	0.1622	0.513
knn	K Neighbors Classifier	0.6767	0.6692	0.8069	0.7540	0.7767	0.1752	0.1838	0.121
ada	Ada Boost Classifier	0.6716	0.6935	0.8069	0.7496	0.7768	0.1551	0.1589	0.109
gbc	Gradient Boosting Classifier	0.6716	0.6774	0.8207	0.7445	0.7802	0.1312	0.1337	0.114
dt	Decision Tree Classifier	0.6155	0.5465	0.7103	0.7374	0.7223	0.0924	0.0924	0.018
qda	Quadratic Discriminant Analysis	0.5491	0.7156	0.4103	0.9206	0.5600	0.2132	0.2996	0.017
nb	Naive Bayes	0.5344	0.7265	0.3690	0.9488	0.5231	0.2161	0.3171	0.018

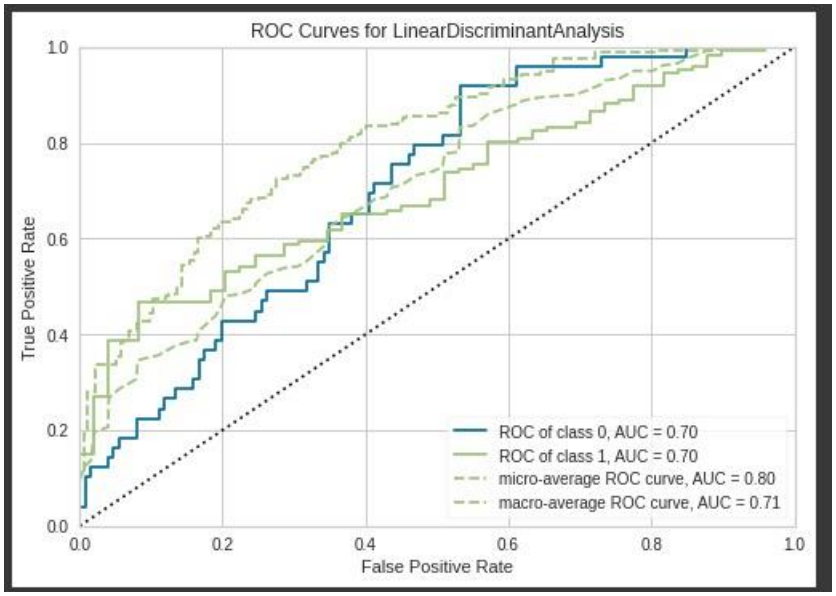
Output Scores Comparison of all the models.

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.7195	0.6674	0.9828	0.7215	0.8321	0.0889	0.1602
1	0.7195	0.6307	0.9138	0.7465	0.8217	0.1947	0.2187
2	0.6829	0.6853	0.9310	0.7105	0.8060	0.0184	0.0251
3	0.7284	0.7804	0.9828	0.7308	0.8382	0.0945	0.1665
4	0.7160	0.8096	0.9828	0.7215	0.8321	0.0362	0.0762
Mean	0.7133	0.7147	0.9586	0.7262	0.8260	0.0865	0.1293
Std	0.0157	0.0685	0.0301	0.0120	0.0114	0.0616	0.0693

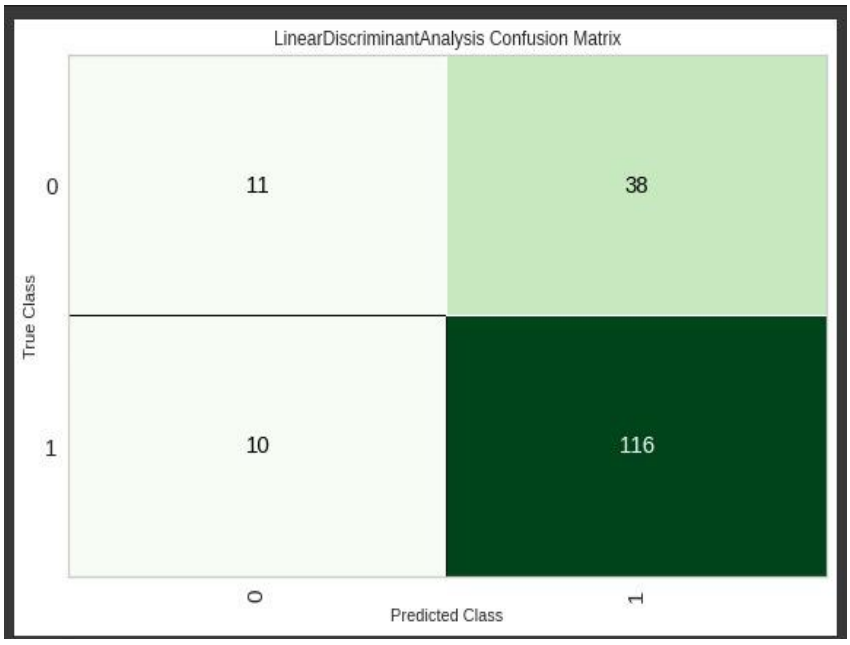
From above results we conclude that LDA gives best accuracy and therefore, we run K-Fold validation on it.

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.5854	0.4943	0.7586	0.6875	0.7213	-0.0806	-0.0821
1	0.7317	0.7845	0.9655	0.7368	0.8358	0.1694	0.2309
2	0.6829	0.6839	0.7931	0.7667	0.7797	0.2150	0.2154
3	0.6829	0.6149	0.8966	0.7222	0.8000	0.0763	0.0879
4	0.7073	0.5718	0.8966	0.7429	0.8125	0.1717	0.1887
5	0.7073	0.7902	0.9310	0.7297	0.8182	0.1214	0.1498
6	0.7561	0.8678	0.9655	0.7568	0.8485	0.2679	0.3305
7	0.7561	0.7213	0.8966	0.7879	0.8387	0.3471	0.3596
8	0.7500	0.7335	0.9310	0.7714	0.8438	0.2453	0.2751
9	0.7750	0.8213	1.0000	0.7632	0.8657	0.2437	0.3725
Mean	0.7135	0.7084	0.9034	0.7465	0.8164	0.1777	0.2128
Std	0.0524	0.1119	0.0720	0.0275	0.0397	0.1129	0.1312

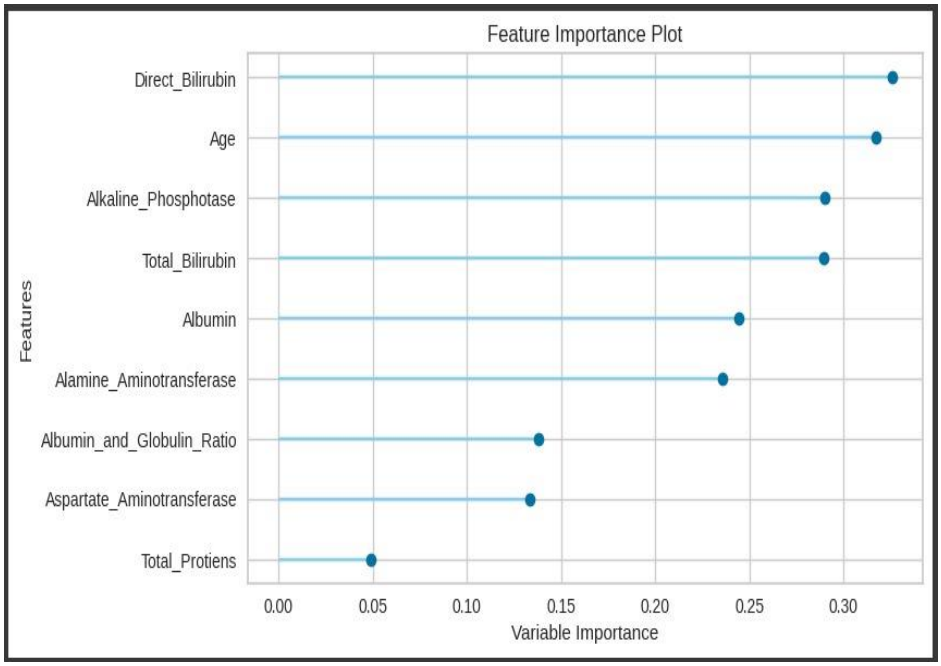
Running 9-Fold Validation gives 77.5% accuracy.



Above is the corresponding ROC Curve for our LDA model



Confusion Matrix.



Feature Importance.

10. ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

No Medical Expertise Required: You don't need to have any knowledge of medical science and liver diseases to predict the liver disease using this application. All we need is Medical Reports which would give use the results of our prediction

High Accuracy: The system predicts the results with accuracy of 100% for this data set. While the accuracy might vary but it is still trustworthy at a large scale.

Immediate Results: The results here are predicted instantaneously thus reducing the time of diagnosis.

Additional Analytics: Based on the analytics we can analyse which patients are most likely to suffer from liver disease in the near future and based on the patient details we will make decisions to cure them.

Scalability: The Machine Learning Model is yet to be integrated with a Web/Mobile Application, thus limiting the number of users and reducing the accessibility.

DISADVANTAGES:

Disease Specific Models: There isn't a disease-specific model that will help in early predictions of different kinds of liver diseases.

Limited Diagnosis: The Final result is binary (whether the patient has 'Risk' or 'No Risk'), thus limiting the scope of diagnosis.

Limited Scope: The web-application is built in flask which is a micro-framework and it is only suitable for small-scale applications. We may not be able to divide a single project into multiple small applications, thus adding new features to the application could be a hassle.

CONCLUSION

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.

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 the model was built by applying various machine learning algorithms. The model was integrated with a flask-based web application.

FUTURE SCOPE

In this project the patients are only classified on the basis of whether they have a risk of Liver Disease or not. There isn't any insight on the exact underlying disease. In the future we could include disease-specific models that will help in early detection of different diseases. The final result is binary, in the future we could include different 'Risk' Categories to further classify patients on the basis of severity.

Rather than using a micro-framework like Flask we could include web-frameworks like Django for our web application. As we would be able to build multiple different applications within the same project in Django.