Statistical Machine Learning Approaches to Liver Disease Prediction

Professional Readiness For Innovation, Employability and Entrepreneurship

IBM-Project-27429 1660056106

Team ID : PNT2022TMID03567

Submitted by

GUNTAMADUGU GANESH	**(212219060098)**
SAI PAVAN PIDIKITI	**(212219063005)**
BOBBA SAI MANOJ	**(212219063001)**
GUNAPATI TEJA	**(212219060095)**

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINNERING

IN

ELECRINICS AND COMMUNICATION ENGINEERING

SAVEETHA ENGINEERING COLLEGE

(AUTONOMOUS)

TABLE OF CONTENTS

ABSTRACT :

Data Mining technologies have been widely used in the process of medical diagnosis and prognosis, extensively. These data mining techniques have been used to analyze a colossal amount of medical data. The steep increase in the rate of obesity and an unhealthy lifestyle eventually reflects the likelihood and the frequent occurrence of liver-related diseases in the mass. 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. Since there are pre-existing processes to analyze the patient data and the classifier data, the more important facet here is to predict the same the conclusive result with a higher rate of accuracy. There are 5 distinct phases in this process. First, the min-max algorithm is applied to the original liver patient data set that could be collected from the UCI repository. In the second phase, significant attributes are demarcated by the use of PSO feature selection. This helps to bring out the subset of critical data, from the whole normalized datasets of liver patients. After this step, the third phase involves the usage of classification algorithms for comparative analysis and categorization. Accuracy Calculation is the fourth phase. It involves the usage of Root Mean Square value and a Root Error value. The fifth phase is the evaluation phase. Depending on the studies, a simple evaluation process is executed to preserve the integrity of a precise result reflection.

CHAPTER 1

INTRODUCTION

1.1 Project Overview

statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers.

The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Machine learning tasks are classified into several broad categories. In supervised learning, the algorithm builds a mathematical model from a set of data that contains both the inputs and the desired outputs. For example, if the task were determining whether an image contained a certain object, the training data for a supervised learning algorithm would include images with and without that object (the input), and each image would have a label (the output) designating whether it contained the object. In special cases, the input may be only partially available, or restricted to special feedback. Classification algorithms and regression algorithms are types of supervised learning.

Classification algorithms are used when the outputs are restricted to a limited set of values. For a classification algorithm that filters emails, the input would be an incoming email, and the output would be the name of the folder in which to file the email. Regression algorithms are named for their continuous outputs, meaning they may have any value within a range. Examples of a continuous value are the temperature, length, or price of an object. In unsupervised learning, the algorithm builds a mathematical model from a set of data which contains only inputs and no desired output labels. Unsupervised learning algorithms are used to find structure in the data, like grouping or clustering of data points. Unsupervised learning can discover patterns in the data, and can group the inputs into categories

In India, delayed diagnosis of diseases is a fundamental problem due to a shortage of medical professionals. A typical scenario, prevalent mostly in rural and somewhat in urban areas is:

1. A patient going to a doctor with certain symptoms.
2. The doctor recommending certain tests like blood test, urine test etc depending on the symptoms.
3. The patient taking the aforementioned tests in an analysis lab.
4. The patient taking the reports back to the reports back to the hospital, where they are examined and the disease is identified.

The aim of this project is to somewhat reduce the time delay caused due to the unnecessary back and forth shuttling between the hospital and the pathology lab.A machine learning algorithm will be trained to predict a liver disease in patients.

1.2 PURPOSE

The purpose of the project is development of web applications of flask frame work for predict the liver disease prediction using machine learning algorithm We designed the frame work and training of different machine learning algorithm of training of dataset

CHAPTER 2

LITERATURE SURVEY

2.1 Title: Neural Networks for Deep Radiotherapy Dose Analysis and Prediction of Liver SBRT Outcomes

Author: Bulat Ibragimov; Diego A. S. Toesca; Yixuan Yuan;

Year: 2019

Description:

Stereotactic body radiation therapy (SBRT) is a relatively novel treatment modality, with little post-treatment prognostic information reported. This study proposes a novel neural network based paradigm for accurate prediction of liver SBRT outcomes. We assembled a database of patients treated with liver SBRT at our institution. Together with a three-dimensional (3-D) dose delivery plans for each SBRT treatment, other variables such as patients' demographics, quantified abdominal anatomy, history of liver comorbidities, other liver-directed therapies, and liver function tests were collected. We developed amulti-path neural network with the convolutional path for 3-D dose plan analysis and fully connected path for other variables analysis, where the network was trained to predict postSBRT survival and local cancer progression. To enhance the network robustness, it was initially pre-trained on a large database of computed tomography images. Following n-fold cross-validation, the network automatically identified patients that are likely to have longer survival or late cancer recurrence, i.e., patients with the positive predicted outcome (PPO) of SBRT, and vice versa, i.e., negative predicted outcome (NPO).

2.2 Title: Predictive Risk Analysis for Liver Transplant Patients - eHealth Model Under National Liver Transplant Program, Uruguay

Author: Parag Chatterjee; Ofelia Noceti; Josemaría Menéndez;

Year: 2020

Description:

Recent years have seen a phenomenal change in healthcare paradigms and data analytics clubbed with computational intelligence has been a key player in this field. One of the main objectives of incorporating computational intelligence in healthcare analytics is to obtain better insights about the patients and proffer more efficient treatment. This work is based on liver transplant patients under the National Liver Transplant Program of Uruguay, considering in detail the health parameters of the patients. Applying computational intelligence helped to separate the cohort into clusters, thereby facilitating the efficient risk-group analysis of the patients assessed under the liver transplantation program with respect to their corresponding health parameters, in a predictive pre-transplant perspective. Also this marks the foundation of Clinical Decision Support Systems in liver transplantation, which act as an assistive tool for the medical personnel in getting a deeper insight to patient health data and thanks to the holistic visualization of the healthcare scenario, also help in choosing a more efficient and personalized treatment strategy.

2.3 Title: Adaptive resonance theory (ARTMAP) for Analysis and Prediction of Survival rate of Patient after Liver Transplantion

Author: Gaurav Soni

Year: 2022

Description:

In both industrialized and developing nations, chronic liver disease affects a significant number of individuals (CLD). Excessive drinking, exposure to toxic gases, and the eating of polluted food all contribute to a rise in the number of liver disease sufferers. Doctors have several challenges in treating patients with liver disease since it affects so many vital bodily systems. A liver transplant is a gruelling procedure with a high likelihood of complications after the procedure. The donor and recipient's compatibility is critical to the transplant's success. If a huge patient and donor database could be utilised to precisely match a donor recipient pair, the post-transplant mortality rate might be considerably reduced. Automated medical diagnostic systems often employ classification algorithms. Artificial neural networks (ANNs), a powerful technology, may aid in the discovery of patterns. It ranges from medicine to the arts that they work in. ANN-based dermatological tools and software may be useful in medicine, according to the conclusions of this study. It was used to analyse and study the Artificial Neural Networks, Radial Basis Function and ARTMAP. ILPD was obtained from the UCI machine learning library. It was possible to evaluate the outcomes using these several methods in order to determine if one method was more accurate than another in terms of precision, accuracy, mean absolute error, and other metrics (root MAE).

2.4 Title: Predictive Cardiometabolic Risk Profiling of Patients Using Vascular Age in Liver Transplantation

Author: Parag Chatterjee; Josemaría Menéndez; Ofelia Noceti; Solange Gerona;

Year: 2021

Description:

Liver transplantation is the last therapeutic option in patients with end-stage liver diseases. The adequate clinical management of transplant-patients impacts their vital prognosis and decisions on many occasions are made from the interaction of multiple variables involved in the process. This work is based on the National Liver Transplantation Program in Uruguay. We performed predictive analysis of cardiometabolic diseases on the transplanted cohort between 2014 and 2019, considering vascular age as a key factor. This aims at classification of the cohort based on the vascular age of the evaluated patients before transplantation for risk-profiling. Predicted high-risk group of the patients showed substantial deterioration of post-transplant health-conditions, including higher mortality rate. In our knowledge, this is the first study in Latin America incorporating vascular age toward predictive analysis of cardio metabolic risk factors in liver transplantations. Predictive risk-modeling using vascular age in a pre-transplantation scenario provides significant opportunity for early prediction of post-transplant risk factors, leading to efficient treatment with anticipation.

2.5 Title: Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques

Author: Sateesh Ambesange; Vijayalaxmi A; Rashmi Uppin**;**

Year: 2021

Description:

Liver disease is a prominent disease other than heart attack, which is taking a lot of lives. Since most of the time liver disease is detected at a later stage leading to death. Number of liver patients is increasing because of several reasons like over consumption of alcohol, breathing in injurious gas, consuming polluted water and so on which can affect health parameters. Using a machine learning prediction models, liver diseases can be predicted using those health parameters in early stages. In this work to build the machine-learning model, Indian Liver Patient Dataset (ILPD) hosted at UCI.edu [1] is used, which is based on Indian patient and Random Forest (RF) algorithm is used to predict the disease with different preprocessing techniques. Data set is checked for skewness, outliers and imbalance using univariate and bivariate analysis and then suitable algorithms used to remove outliers and various oversampling and under sampling techniques are used to balance the data. Further refinement of model is done through hyper parameter tuning using grid search and feature selection. The final model provides 100% accuracy and also good score across different metrics.

2.6 Title: Irregular Respiratory Motion Compensation for Liver Contrast-Enhanced Ultrasound via Transport-Based Motion Estimation

Author: Peng Wan; Fang Chen; Wei Shao; Chunrui Liu

Year: 2020

Description:

Contrast-enhanced ultrasound (CEUS) imaging has been widely applied for the detection and characterization of focal liver lesions (FLLs), for its ability to visualize the blood flow in real time. However, cyclic liver motion poses a great challenge to the recovery of perfusion curves as well as quantitative kinetic parameters estimation. Recently, a few gating methods have been proposed to eliminate unexpected intensity fluctuations by the breathing phase estimation, with the assumption that each breathing phase corresponds to a specific lesion position strictly. While practical liver motion tends to be irregular due to changes in the patient's underlying physiologic status, that is, the same phase might not correspond to the same position. To tackle the challenge of motion irregularity, we present a novel motion estimation-based respiratory compensation method, named RCME, which first estimates salient motion information through the framework of optimal transport (OT) by jointly modeling pixel intensity as well as their locations and then employs sparse subspace clustering (SSC) to identify the subset of frames acquired at the same position. Our proposed method is evaluated on 15 clinical CEUS sequences in both qualitative and quantitative manners. Experimental results demonstrate good performance on irregular liver motion compensation.

2.7 Title: Prediction of Graft Dysfunction in Pediatric Liver Transplantation by Logistic Regression

Author: Krasimira Prodanova; Yordanka Uzunova

Year: 2020

Description:

Liver transplantation (LT) is the last therapeutic option for children with acute and chronic end-stage liver disease. Nowadays survival rates are over 90-95% and 70% within one year and five years post-transplantation, respectively. The main complications in the postoperative period are related to the function of the graft. The graft dysfunction early after LT is an important cause of morbidity and mortality. Numerous factors can affect the function of the graft after pediatric LT. Therefore our aim is to identify the risk factors in order to help prevent the graft failure. In the present work, by means of univariate and multivariate logistic regression analysis, the probabilities of graft dysfunction in the early postoperative period after LT are estimated. As predictors in the constructed logistic models the following parameters have been analyzed: levels of bilirubin, sodium, creatinine, international normalized ratio (INR) in blood plasma, post-transplant MELD score (Model for End-Stage Liver Disease) and cold ischemia time on the 1 st , 2 nd , 3 rd , 5 th , 7 th and 10 th postoperative day. The models were based on 31 patient's data obtained at the University Hospital "Lozenets" - Sofia, Bulgaria.

2.8 Title: Towards Chronic Liver Dysfunction Self-monitoring: a Proof-of-Concept Study

Author: Danila Germanese; Sara Colantonio; Mario D'Acunto; Maurizia Brunetto;

Year: 2021

Description:

The liver is our very own chemical processing plant as it plays a vital role in maintaining the body's metabolic balance. Liver's health is assessed by a group of clinical tests (such as blood tests, ultrasonographic imaging, liver biopsy) most of which are invasive and burdensome for the patients. In the setting of severely scarred liver, toxic substances, such as ammonia, have fewer opportunities to be detoxified. Accumulation of ammonia in the systemic circulation and in the brain may result in Hepatic Encephalopathy (HE), a spectrum of neuropsychiatric abnormalities which entails changes in consciousness, intellectual functions, behavior. Minimal HE has attracted increasing attention, as it does not cause detectable changes in personality or behaviour, but the complex and sustained attention is impaired. Hence, it can be detected only by specific but biased, time-consuming and burdensome examinations, such as blood ammonia levels assessment and neuro-psychological tests. The obstrusivity of the majority of the liver function clinical tests, and, in case of minimal HE, the lack of reliable examinations, are encouraging the scientific community to look for alternative diagnostic methods. For this purpose, the exploitation of a non-invasive technique such as breath analysis, to identify chronic liver disease, discriminate among its degree of severity and detect the onset of HE, could be a step forward for clinical diagnosis.

2.9 Title: Comparative Analysis of Liver diseases by using Machine Learning Techniques

Author: D. Palanivel Rajan; K S Kannan; P. Divya;

Year: 2022

Description:

In a human body function of the liver is important. Many persons are suffering from liver disease, but they don't know it. The identification of liver diseases in the early stage helps a patient get better treatment. If it is not diagnosed earlier, it may lead to various health issues. To solve these issues, physicians need to examine whether the patient has been affected by liver disease or not, based on the multiple parameters. In this paper, we classify the patients who have liver disease or not by using different machine learning algorithms by comparing the performance factors and predicting the better result. The liver dataset is retrieved from the Kaggle dataset.

2.10 Title: Analysis of Liver Disorder Using Classification Techniques: A Survey

Author: Sushmit Pahari; Dilip Kumar Choubey

Year: 2021

Description:

Classification is an effective and widespread techniques used in many medical fields like liver diagnosis of patients. The prime goal of this research article is to the summarize many popular existing techniques for the classification of liver disorder. The techniques have been analysed based on their advantages, issues as well as accuracy. So, on the basis of issues, a new hybrid classification technique may be designed. This research article has been concluded with some of the recent research article of future directions.

2.1 EXISTING SYSTEM

Using machine learning algorithms to predict disease is made possible by increasing access to hidden attributes in medical data sets. Various kinds of data sets, such as blood panels with liver function tests, histologically stained slide images, and the presence of specific molecular markers in blood or tissue samples, have been used to train classifier algorithms to predict liver disease with good accuracy. The ML methods described in previous studies have been evaluated for accuracy by a combination of confusion matrix, receiver operating characteristic under area under curve, and k-fold cross-validation. Singh et al. designed software based on classification algorithms (including logistic regression, random forest, and naive Bayes) to predict the risk of liver disease from a data set with liver function test results.

Phan and Chan et al. demonstrated that a convolutional neural network (CNN) model predicted liver cancer in subjects with hepatitis with an accuracy of 0.980 . The ANN model has been used to predict liver cancer in patients with type 2 diabetes . Neural network ML methods can help differentiate between types of liver cancers when applied to imaging data sets. Neural network algorithms have even been trained to predict a patient’s survival after liver tumor removal using a data set containing images of processed and stained tissue from biopsies . ML methods can facilitate the diagnosis of many diseases in clinical settings if trained and tested thoroughly.

This study aimed to

(i) compute missing data using the MICE algorithm.

(ii) determine variable selection using eigen decomposition of a data matrix by PCA and to rank the important variables using the Gini index

(iii) compare among several statistical learning methods the ability to predict binary classifications of liver disease

(iv) use the synthetic minority oversampling technique (SMOTE) to oversample minority class to regulate overfitting

(iv) obtain confusion matrices for comparing actual classes with predictive classes

(v) compare several ML approaches to assess a better performance of liver disease diagnosis

(vi) evaluate receiver operating characteristic (ROC) curves for determining the diagnostic ability of binary classification of liver disease.

2.2 REFERENCES

[1] F. C. Robey, D. R. Fuhrman, and S. I. Krich, “Array calibration utilizing clutter scattering,” in Proc. IEEE 7th SP Workshop on Statistical Signal and Array Processing, 1994, pp. 317–320.

[2] H. Steyskal and J. S. Herd, “Mutual coupling compensation in small array antennas,” in Proc. IEEE Aerospace Conf., 2002, vol. 2, pp. 827–831.

[3] R. Klemm, Space-Time Adaptive Processing Principles and Applications. London, U.K.: Inst. Elect. Eng., 1998.

[4] J. C. Spall, “Multivariate stochastic approximation using a simultaneous perturbation gradient approximation,” IEEE Trans. Autom. Control, vol. 37, pp. 332–341, 1992.

[5] J. C. Spall, Introduction to Stochastic Search and Optimization. New York, NY, USA: Wiley, 2003.

[6] M. L. Pugh and P. A. Zulch, “RLSTAP algorithm development tool for analysis of advanced signal processing techniques,” in Proc. 29th Asilomar Conf. on Signals, Systems and Computers, 1996, vol. 2, pp. 1178–1182.

[7] Radar Analysis/Simulation Tool for Khoros (RASTK). CAESoft Corp, Rockwall, TX, USA.

[8] A. D. Belegundu and T. R. Chandrupatla, Optimization Concepts and Applications in Engineering. Cambridge, U.K.: Cambridge Univ. Press, 2011, pp. 104–106.

[9] D. Wilson, “Asymptotic Properties of Nearest Neighbor Rules Using Edited Data,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 2(3), pp. 408-421, 1972.

[10] J. Laurikkala, “Improving identification of difficult small classes by balancing class distribution,” Proceedings of the 8th Conference on Artificial Intelligence in Medicine in Europe, pp. 63-66, 2001

[11] P. E. Hart, “The condensed nearest neighbor rule,” IEEE Transactions on Information Theory, vol. 14(3), pp. 515-516, 1968.

[12] Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J & Napolitano, A. “RUSBoost: A hybrid approach to alleviating class imbalance.” IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 40.1 (2010): 185-1

2.3 PROBLEM STATEMENT

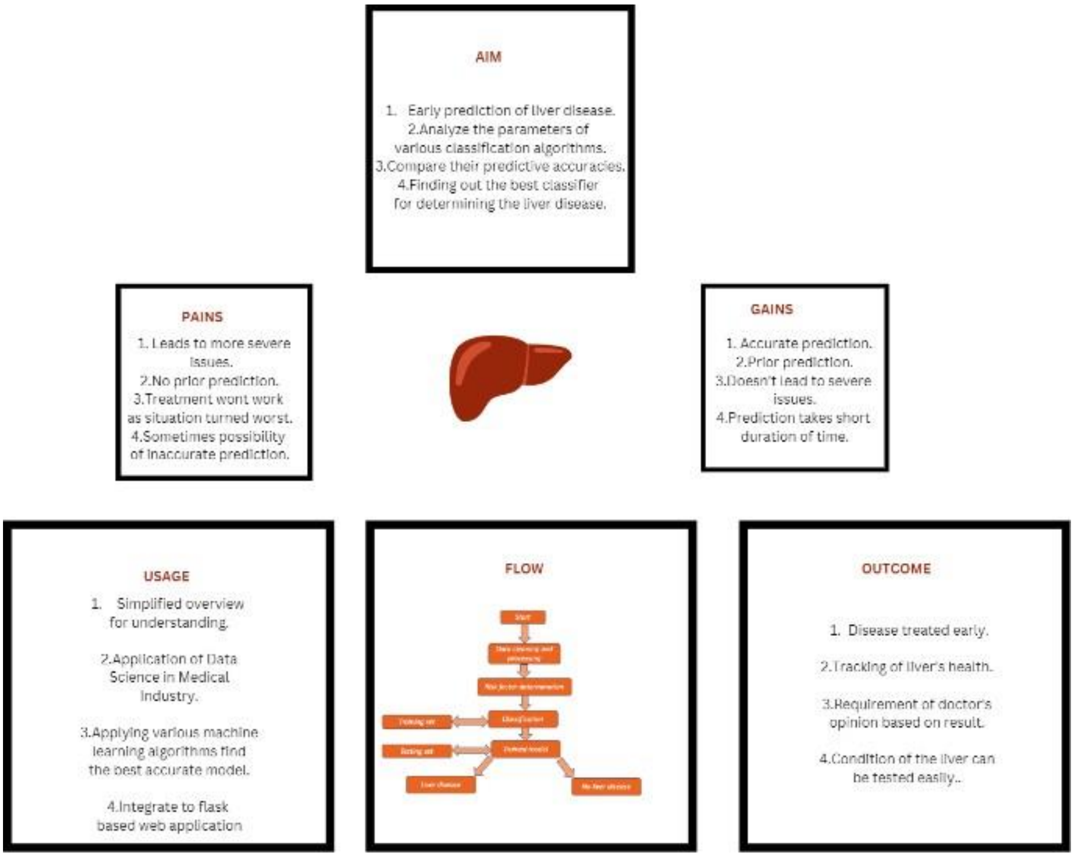
The Problem Statement Is Formally Defined As:

‘Given A Dataset Containing Various Attributes Of 584 Indian Patients, Use The Features Available In The Dataset And Define A Supervised Classification Algorithm Which Can Identify Whether A Person Is Suffering From Liver Disease Or Not. This Data Set Contains 416 Liver Patient Records And 167 Non- Liver Patient Records. The Data Set Was Collected From North East Of Andhra Pradesh, India. This Data Set Contains 441 Male Patient Records And 142 Female Patient Records. Any Patient Whose Age Exceeded 89 Is Listed As Being Of Age "90".

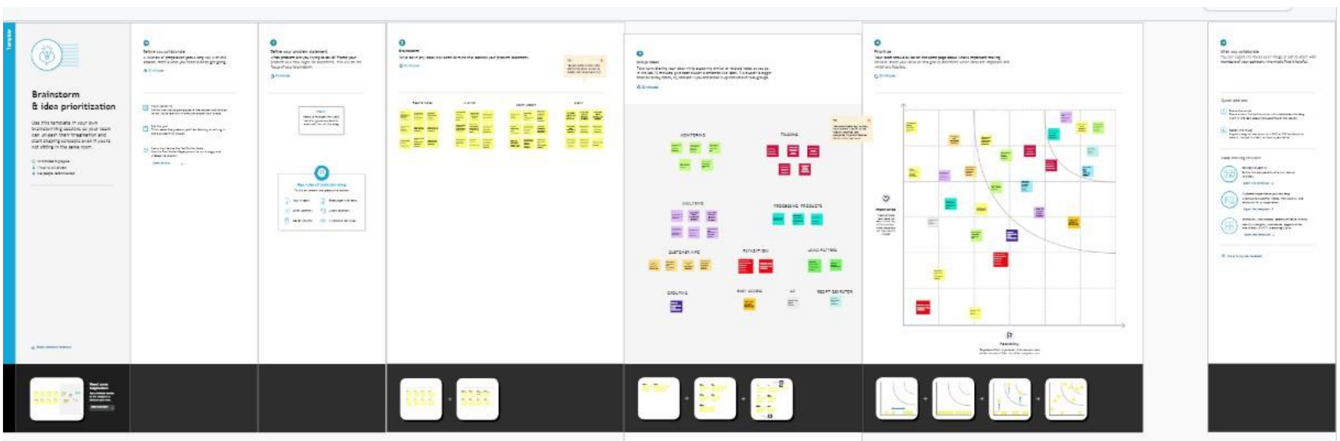
CHAPTER 3

IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION & BRAINSTORMING



3.3 PROPOSED SOLUTION

Liver diseases avert the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises. Discovering the existence of liver disease at an early stage is a complex task for the doctors. Very satisfied, happier, healthier due to earlier and accurate prediction of liver diseases.

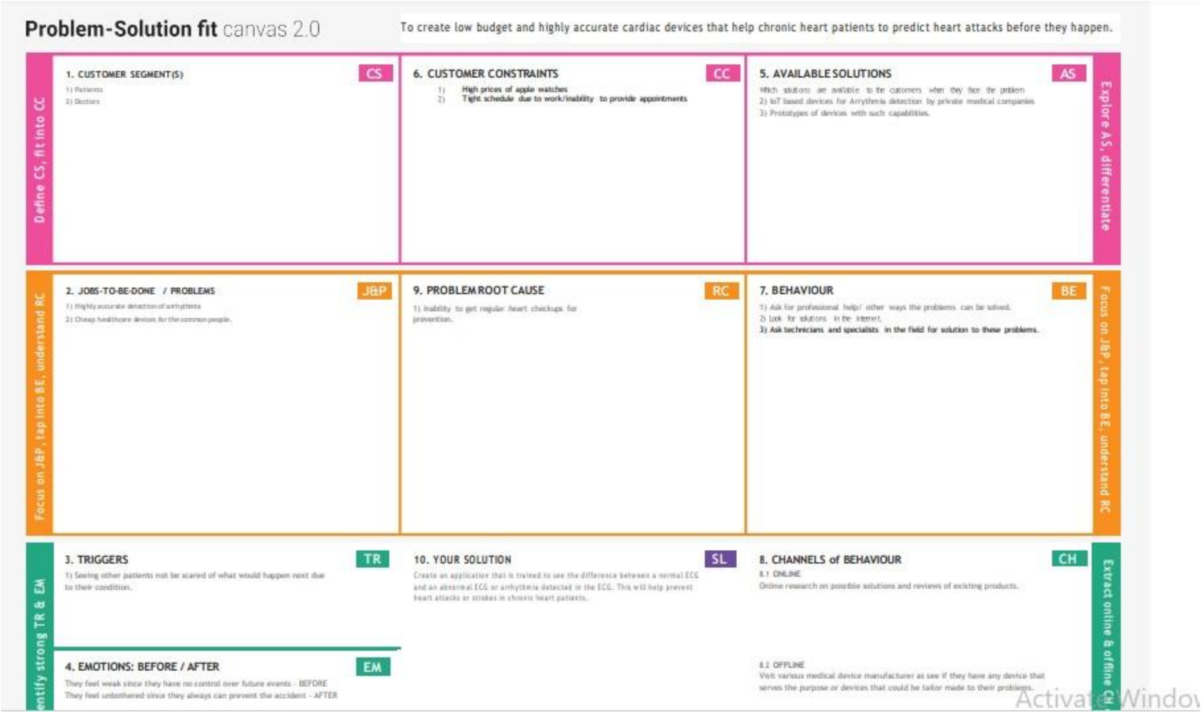
Use our project to examine data from liver patients concentrating on relationships between a key list of liver enzymes, proteins, age and gender using them to try and predict the likeliness of liver disease. Here we are building a model by applying various machine learning algorithms find the best accurate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

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

Phan and Chan et al. demonstrated that a convolutional neural network (CNN) model predicted liver cancer in subjects with hepatitis with an accuracy of 0.980 . The ANN model has been used to predict liver cancer in patients with type 2 diabetes . Neural network ML methods can help differentiate between types of liver cancers when applied to imaging data sets.

Liver diseases avert the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises.Discovering the existence of liver disease at an early stage is a complex task for the doctors.

3.4 PROBLEM SOLUTION FIT



CHAPTER 4

REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form present in liver disease prediction website
FR-2	User Confirmation	Confirmation via Email. An email will be sent to user on registered email for confirmation.
FR-3	Patient details like age, proteins, enzyme etc	Collected via website input fields
FR-4	Prediction	Based on liver enzymes, proteins, age and gender using them to try and predict the likeliness of liver disease.
FR-5	Retrieval(Database)	Whenever we need user data we can get it from database.
FR-6	Internet connection	As it is web based application user must have internet connectivity to access it.

4.2 NON-FUNCTIONAL REQUIREMENTS

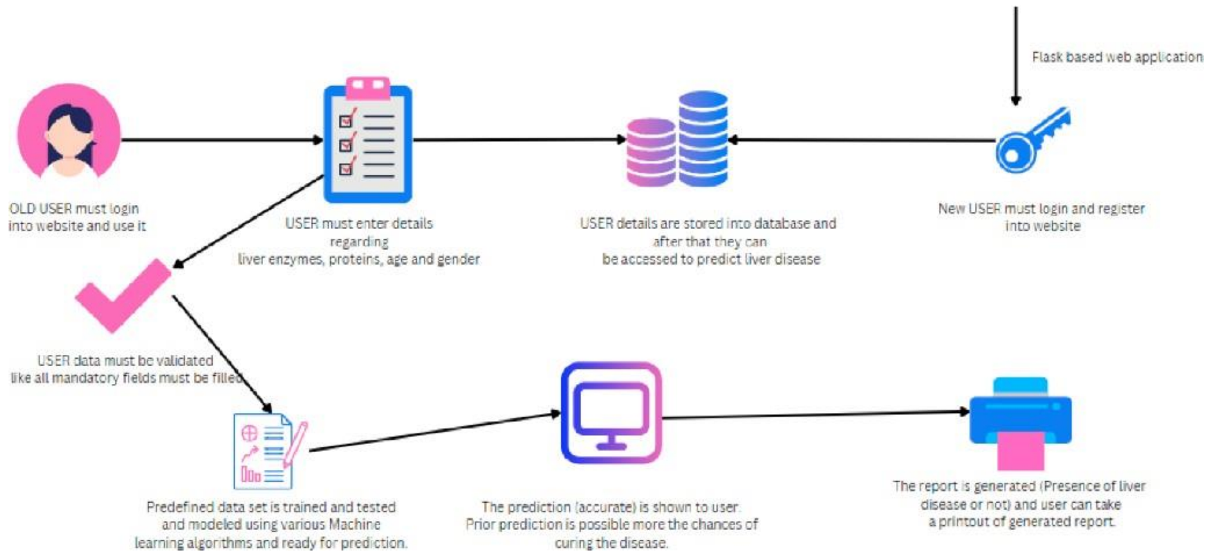
A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data centers and leaves the system, what changes the information, and where data is stored.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Easily usable as it is web based application people can just type URL to get the web page.
NFR-2	**Security**	Highly secured by getting SSL certificate and using block chain technologies.
NFR-3	**Reliability**	Consistently good in quality and is trusted because we are securely storing the data in database and no one can steal it and bloack chain based security is also used.
NFR-4	**Performance**	Performance is high as we are using various Machine learning algorithms find the best accurate model. And integrate to flask based web application. Therefore performance rich.
NFR-5	**Availability**	Application will be active once user registers with it and can be used any time needed.
NFR-6	**Scalability**	Highly scalable as we are going to use IBM Cloud.

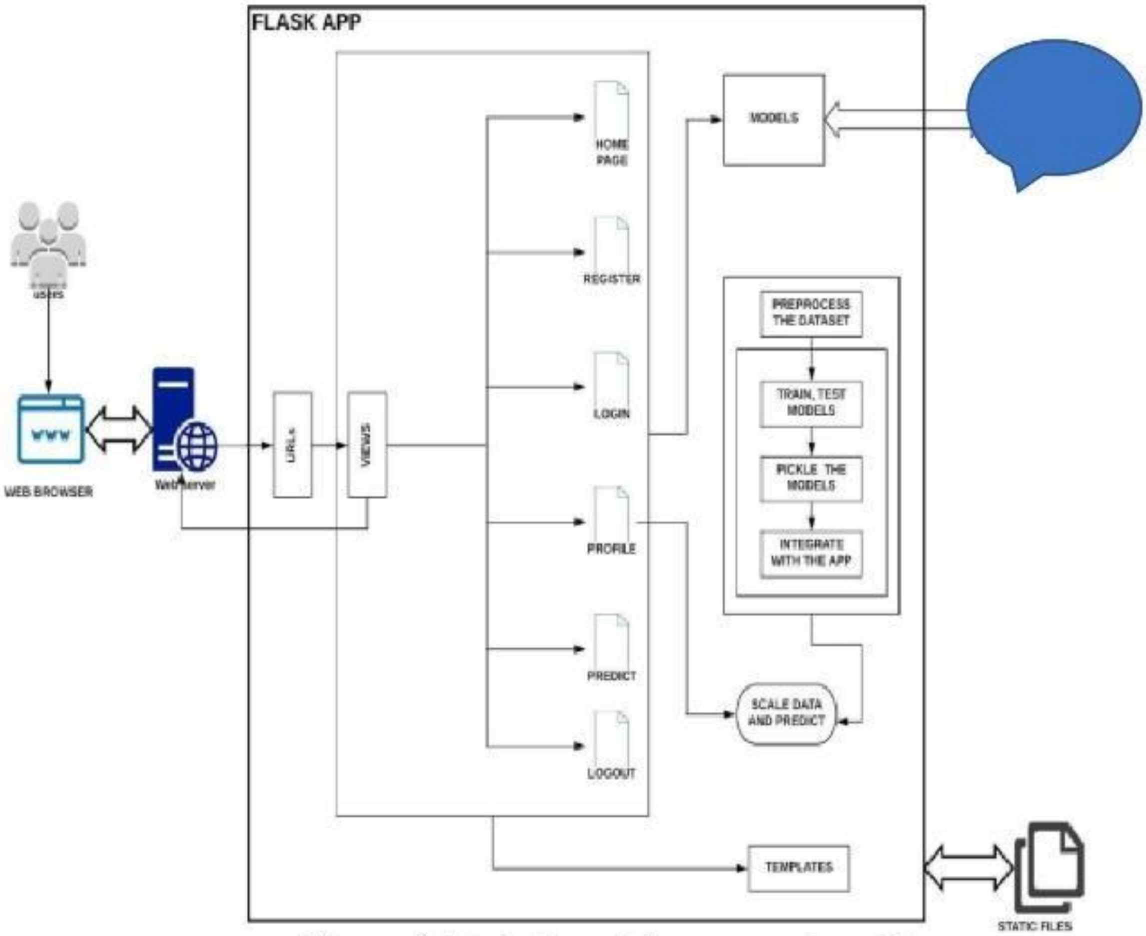
CHAPTER 5

PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS



5.2 SOLUTION & TECHNICAL ARCHITECTURE



5.3 USER STORIES

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	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.	High	Kavuri Swathi Karumudi Meghana Biridepalli Mounisha Jasthi Jyothi
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	High	Kavuri Swathi Karumudi Meghana Biridepalli Mounisha Jasthi Jyothi
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email & password	High	Kavuri Swathi Karumudi Meghana Biridepalli Mounisha Jasthi Jyothi
Sprint-2	Input Necessary Details	USN-4	As a user, I can give Input Details to Predict Likeliness of Liver Disease.	High	Kavuri Swathi Karumudi Meghana Biridepalli Mounisha Jasthi Jyothi
Sprint-2	Data Pre-Processing	USN-5	Transform raw data into suitable format for prediction.	High	Kavuri Swathi Karumudi Meghana Biridepalli Mounisha Jasthi Jyothi
Sprint-3	Prediction of Liver Disease	USN-6	As a user, I can predict Liver Disease using machine learning model.	High	Kavuri Swathi Karumudi Meghana Biridepalli Mounisha Jasthi Jyothi
Sprint-3		USN-7	As a user, I can get accurate prediction of liver disease.	Medium	Kavuri Swathi Karumudi Meghana Biridepalli Mounisha Jasthi Jyothi
Sprint-4	Review	USN-8	As a user, I can give feedback of the application.	High	Kavuri Swathi Karumudi Meghana Biridepalli Mounisha Jasthi Jyothi

CHAPTER 6

PROJECTS PLANNING & SCHEDULING

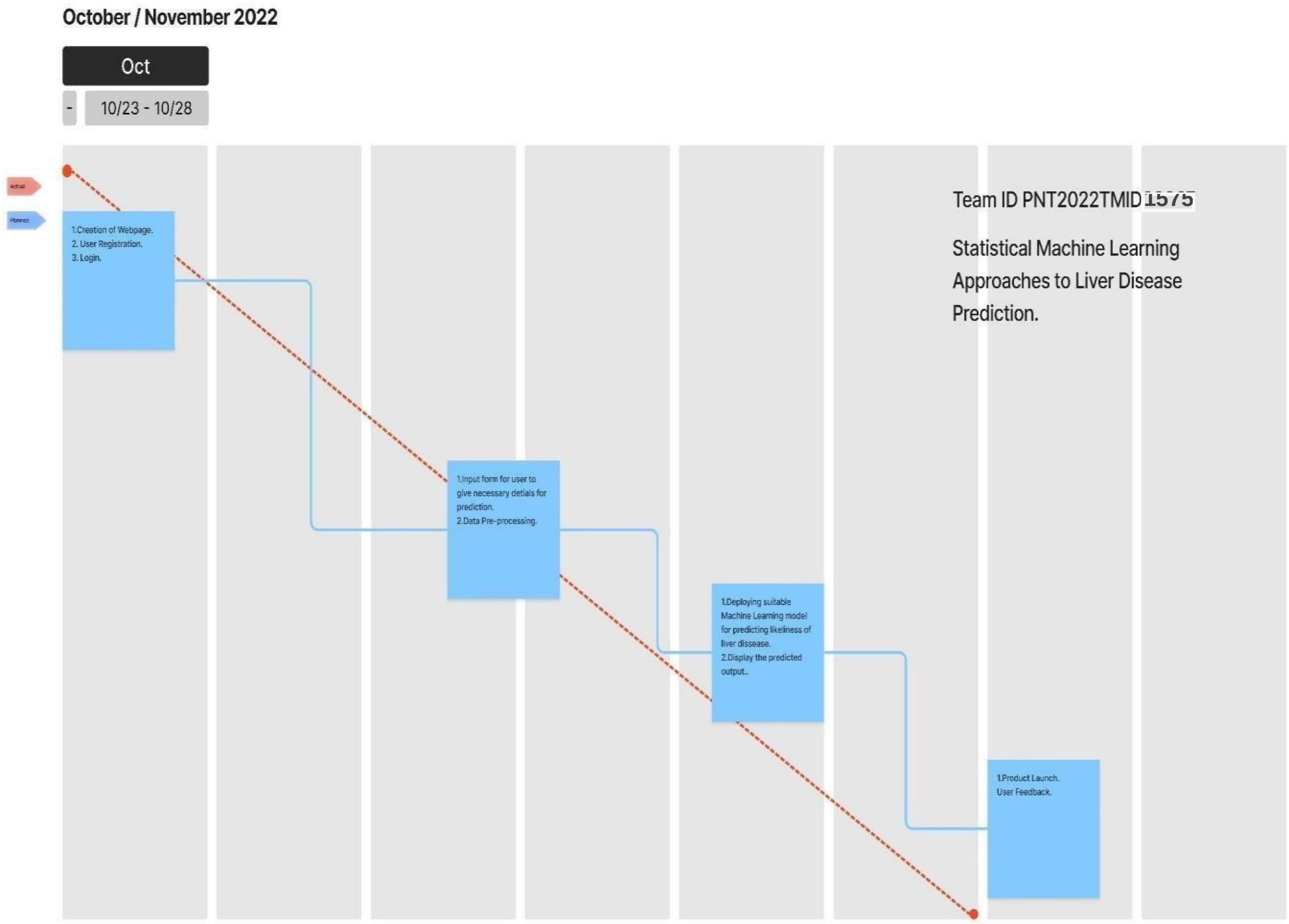
6.1 SPRINT PLANNING & ESTIMATION

Sprint	Duration	Sprint Start Date	Sprint End Date (Planned)	Sprint Release Date (Actual)
Sprint-1	6 Days	24 Oct 2022	29 Oct 2022	29 Oct 2022
Sprint-2	6 Days	31 Oct 2022	05 Nov 2022	
Sprint-3	6 Days	07 Nov 2022	12 Nov 2022	
Sprint-4	6 Days	14 Nov 2022	19 Nov 2022	

6.2 SPRINT DELIVERY SCHEDULE

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Priority
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	High
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	High
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email & password	High
Sprint-2	Input Necessary Details	USN-4	As a user, I can give Input Details to Predict Likeliness of Liver Disease.	High
Sprint-2	Data Pre-Processing	USN-5	Transform raw data into suitable format for prediction.	High
Sprint-3	Prediction of Liver Disease	USN-6	As a user, I can predict Liver Disease using machine learning model.	High
Sprint-3		USN-7	As a user, I can get accurate prediction of liver disease.	Medium
Sprint-4	Review	USN-8	As a user, I can give feedback of the application.	High

6.3 REPORTS FROM JIRA



CHAPTER 7

CODING & SOLUTIONING

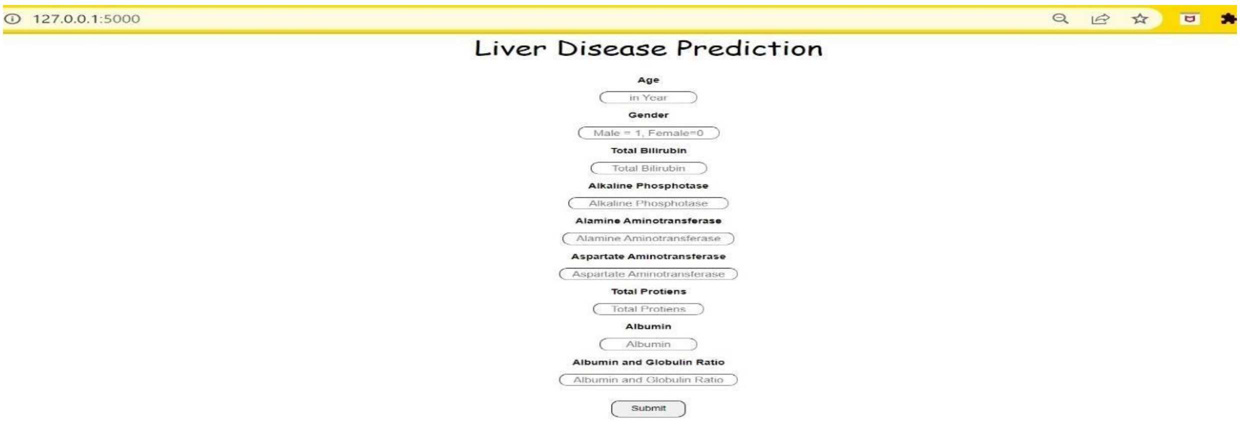
7.1 FEATURE 1

Index.html(code):

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="UTF-8">
    <title>Liver Prediction Model</title>

</head>
<body>
    <div class="container">
        <h2 class='container-heading'><span class="heading_font">Liver Disease
Prediction</span></h2>
    </div>

    <div class="ml-container">
        <form action="{{ url_for('predict') }}" method="POST">
            <br>
            <h3>Age</h3>
            <input id="first" name="Age" placeholder="in Year"
required="required">
            <br>
            <h3>Gender</h3>
            <input id="second" name="Gender" placeholder="Male = 1, Female=0"
required="required">
            <br>
            <h3>Total Bilirubin</h3>
            <input id="third" name="Total_Bilirubin" placeholder="Total
Bilirubin" required="required">
            <br>
            <h3>Alkaline Phosphotase</h3>
            <input id="fourth" name="Alkaline_Phosphotase"
placeholder="Alkaline Phosphotase" required="required">
            <br>
            <h3>Alamine Aminotransferase</h3>
            <input id="fifth" name="Alamine_Aminotransferase"
placeholder="Alamine Aminotransferase" required="required">
            <br>
            <h3>Aspartate Aminotransferase</h3>
            <input id="sixth" name="Aspartate_Aminotransferase"
placeholder="Aspartate Aminotransferase" required="required">
```

```
            <br>
            <h3>Total Protiens</h3>
            <input id="seventh" name="Total_Protiens" placeholder="Total 
Protiens" required="required">
            <br>
            <h3>Albumin</h3>
            <input id="eight" name="Albumin" placeholder="Albumin" 
required="required">
            <br>
            <h3>Albumin and Globulin Ratio</h3>
            <input id="ninth" name="Albumin_and_Globulin_Ratio" 
placeholder="Albumin and Globulin Ratio" required="required">
            <br>
            <br>
            <br>
            <button id="sub" type="submit ">Submit</button>
            <br>
            <br>
            <br>
            <br>


        </form>
    </div>
</body>
</html>
```

Index.html(output):

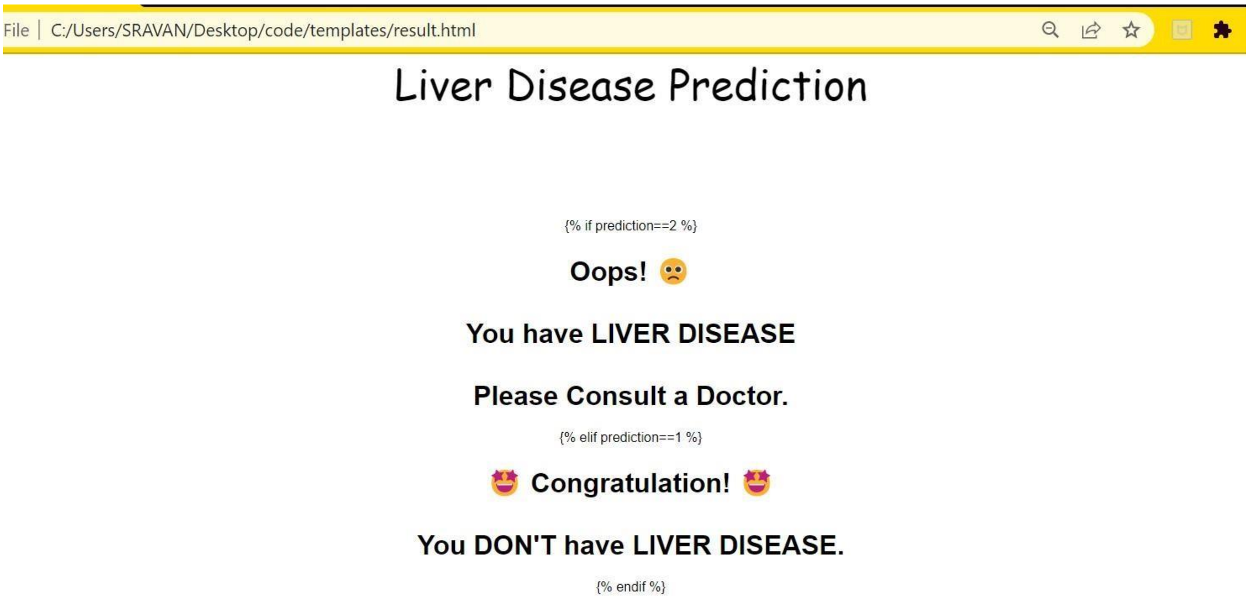


7.2 FEATURE 2

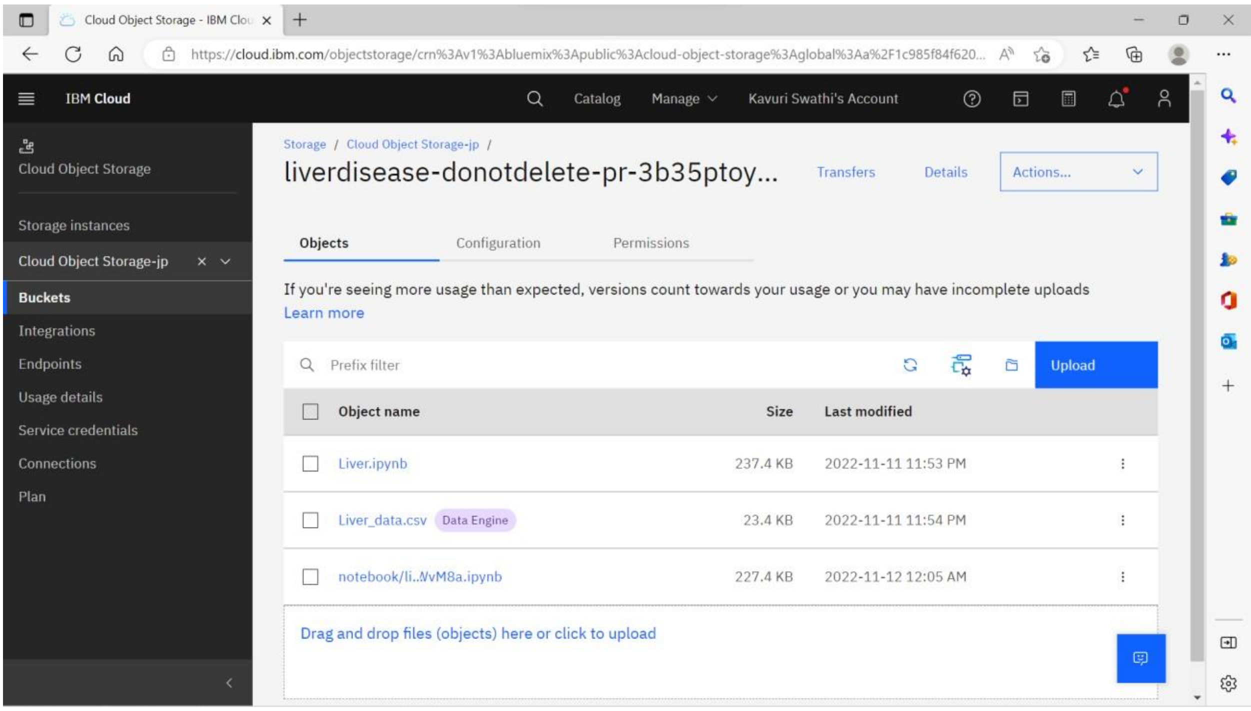
Result.html(code):

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Liver Disease Result</title>
</head>

<body>

    <div class="container">
        <form action="{{ url_for('predict')}}" method="post">
            <h2 class='container-heading'><span class="heading_font">Liver
Disease Prediction</span></h2>

        <br><br><br><br><br><br><br>

        <!-- Result -->
        <div class="results">
            {% if prediction==2 %}
                <h1><span class='danger'>Oops! ☹<br><br>You have LIVER
DISEASE <br><br>Please Consult a Doctor.</span></h1>

            {% elif prediction==1 %}
                <h1><span class='safe'>□ Congratulation! □<br><br>You DON'T
have LIVER DISEASE.</span></h1>

            {% endif %}
        </div>
        </form>

    </div>
    <div>
            <br><br> <br><br><br><br><br><br><br><br>


    </div>
```

Result.html(output):



7.3 DATABASE SCHEMA



CHAPTER 8

TESTING

8. SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

8.1 TYPES OF TESTS

8.1.1 Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

8.1.2 Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

8.1.3 Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

8.1.4 System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

8.1.5 White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

8.1.6 Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

8.2 Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

8.2.1 Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

8.2.2 Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

8.2.3 Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

8.3 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

8.4 Acceptance Testing

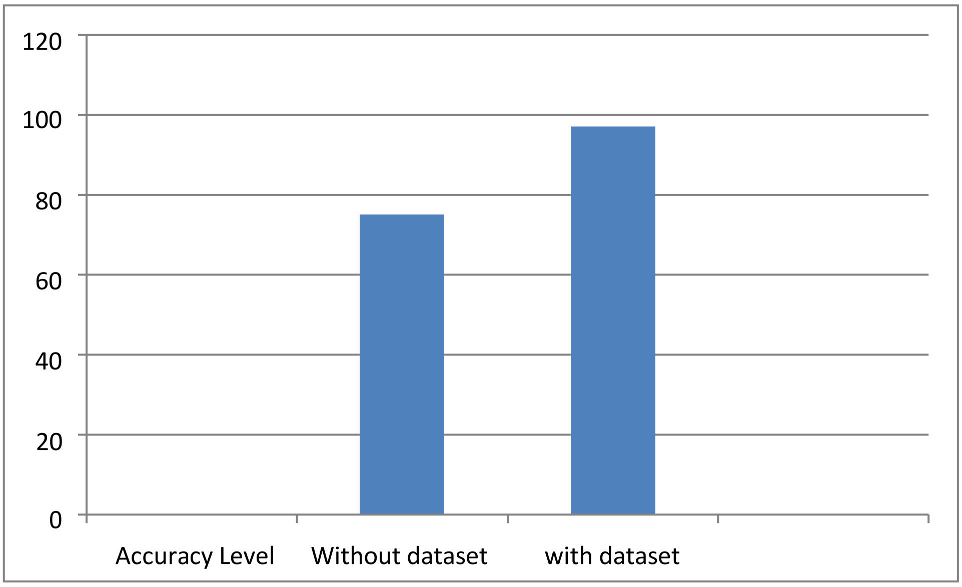
User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

CHAPTER 9

RESULTS

9.1 PERFORMANCE METRICS



CHAPTER 10

10 ADVANTAGE & DISADVANTAGE

Machine learning is a functionality of a system to be able to learn through the extensive usage of examples that pose a set of conditions that can be incorporated as a part of the self improvement process without being coded by a programmer.

The result, thus obtained is then used by the corporate, in order to make actionable inferences for decision making. It has its roots related to data mining and close association with Bayesian predictive modelling.

The data is taken as an input by the machine and the result is formulated as the output. Typical machine learning algorithms are utilized in trying to improve the user experience by providing recommendations using historical data. This would be an opportunistic approach to utilise this unsupervised learning to do the same.

- High accuracy
- High sensitivity
- High prediction rate
- Reduced error of the system

CHAPTER 11

CONCLUSIONS

we designed web based Diseases related to liver and heart are becoming more and more common with time. With continuous technological advancements, these are only going to increase in the future. Although people are becoming more conscious of health nowadays and are joining yoga classes, dance classes; still the sedentary lifestyle and luxuries that are continuously being introduced and enhanced; the problem is going to last long.

So, in such a scenario, our project will be extremely helpful to the society. With the dataset that we used for this project, we got 100 % accuracy for SVM model, and though it might be difficult to get such accuracies with very large datasets, from this projects results, one can clearly conclude that we can predict the risk of liver diseases with accuracy of 90 % or more.

Today almost everybody above the age of 12 years has smartphones with them, and so we can incorporate these solutions into an android app or ios app. Also it can be incorporated into a website and these app and website will be highly beneficial for a large section of society.

CHAPTER 12

FUTURE SCOPES

- In Future Scope We Developed New Machine Learning Algorithm Real Time Implementation Design For Web Technology For Prediction Of Liver Disease Prediction Rate
- New development and frame work of liver disease prediction with reduced error and improve accuracy

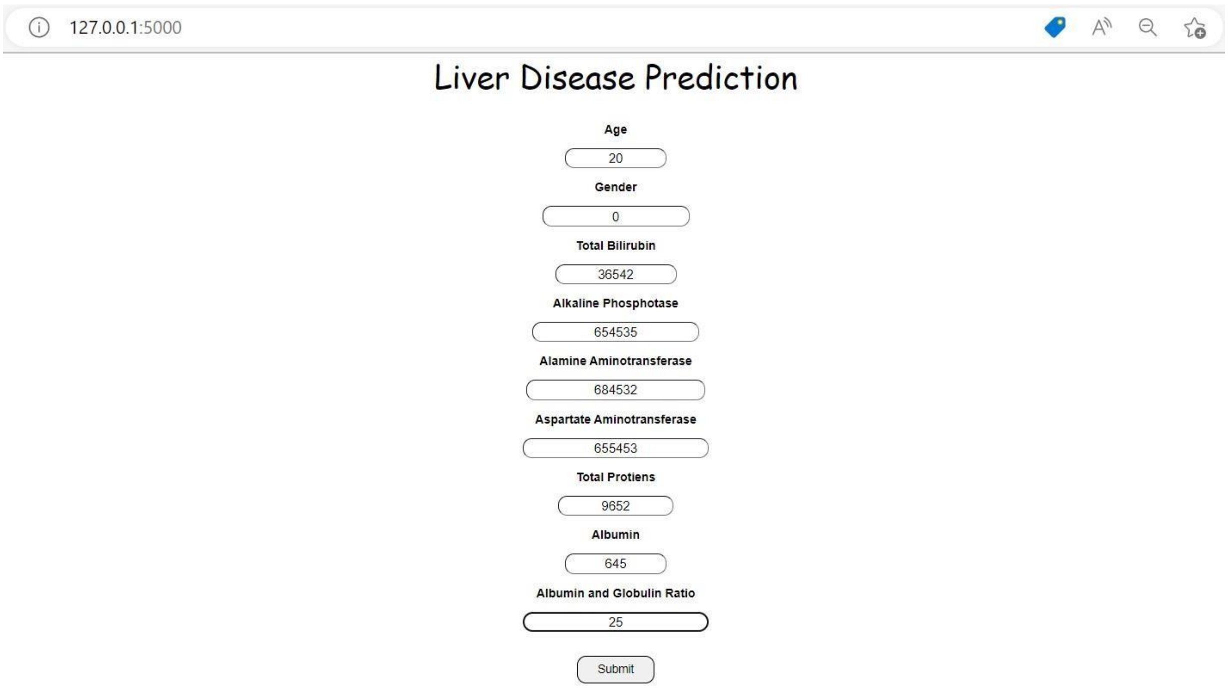
CHAPTER 13

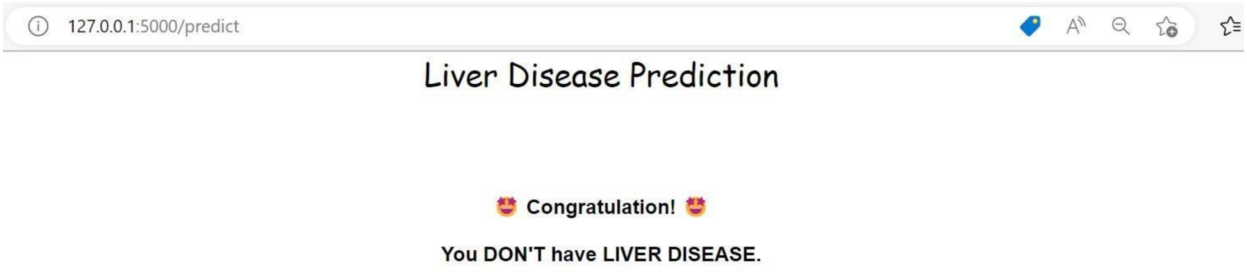
APPENDIX

13.1 SOURCE CODE

```
from flask import Flask, render_template, request

import numpy as np

import pickle

app = Flask(__name__)

model = pickle.load(open('Liver2.pkl', 'rb'))

@app.route('/',methods=['GET'])

def Home():

    return render_template('index.html')

@app.route("/predict", methods=['POST'])

def predict():

    if request.method == 'POST':

        Age = int(request.form['Age'])

        Gender = int(request.form['Gender'])

        Total_Bilirubin = float(request.form['Total_Bilirubin'])

        Alkaline_Phosphotase = int(request.form['Alkaline_Phosphotase'])

        Alamine_Aminotransferase                                    =
int(request.form['Alamine_Aminotransferase'])

        Aspartate_Aminotransferase                                  =
int(request.form['Aspartate_Aminotransferase'])
```

```
    Total_Protiens = float(request.form['Total_Protiens'])

    Albumin = float(request.form['Albumin'])

    Albumin_and_Globulin_Ratio = float(request.form['Albumin_and_Globulin_Ratio'])

    values = np.array([[Age,Gender,Total_Bilirubin,Alkaline_Phosphotase,Alamine_Aminotransferase,Aspartate_Aminotransferase,Total_Protiens,Albumin,Albumin_and_Globulin_Ratio]])

    prediction = model.predict(values)

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

if__name__== "__main__":

  app.run(debug=True)
```





13.2 GitHub & Project Demo Link

Git-hub link : https://github.com/IBM-EPBL/IBM-Project-27429-1660056106

Project demo-link: https://youtu.be/msbQBVVifS0