1.INTRODUCTION

1.1 Project Overview:

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

1.2 Purpose:

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

2.LITERATURE SURVEY

2.1 Existing Problem:

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

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

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

Data Mining technologies have been widely used in the process of medical diagnosis and prognosis, extensively. In this project, the patient data sets are analyzed for the predictability of the

subject to have a liver disease based purely on a widely analyzed classification model. This System predict the same conclusive result with a higher rate of accuracy. J48 algorithm is considered to be a better performing algorithm when it comes to feature selection with an accuracy rate of 95.04%.

Various kinds of pressure and unbalanced eating behaviors, along with alcohol inhalation and on-going toxic gases, etc, cause lever disease in patients.For this purpose, the type of data mining algorithms can help medical doctors to diagnose patients in hospital. This paper analyzes meta learning algorithms to classify the Indian liver patient dataset. Adaboost, logitboost, Bagging and Grading meta learning algorithms are applied to this data set. Key role is played by Grading algorithm in shaping enhanced classification accuracy (Correct Classification Rate) of a data set.

2.2 Reference:

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

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

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

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

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

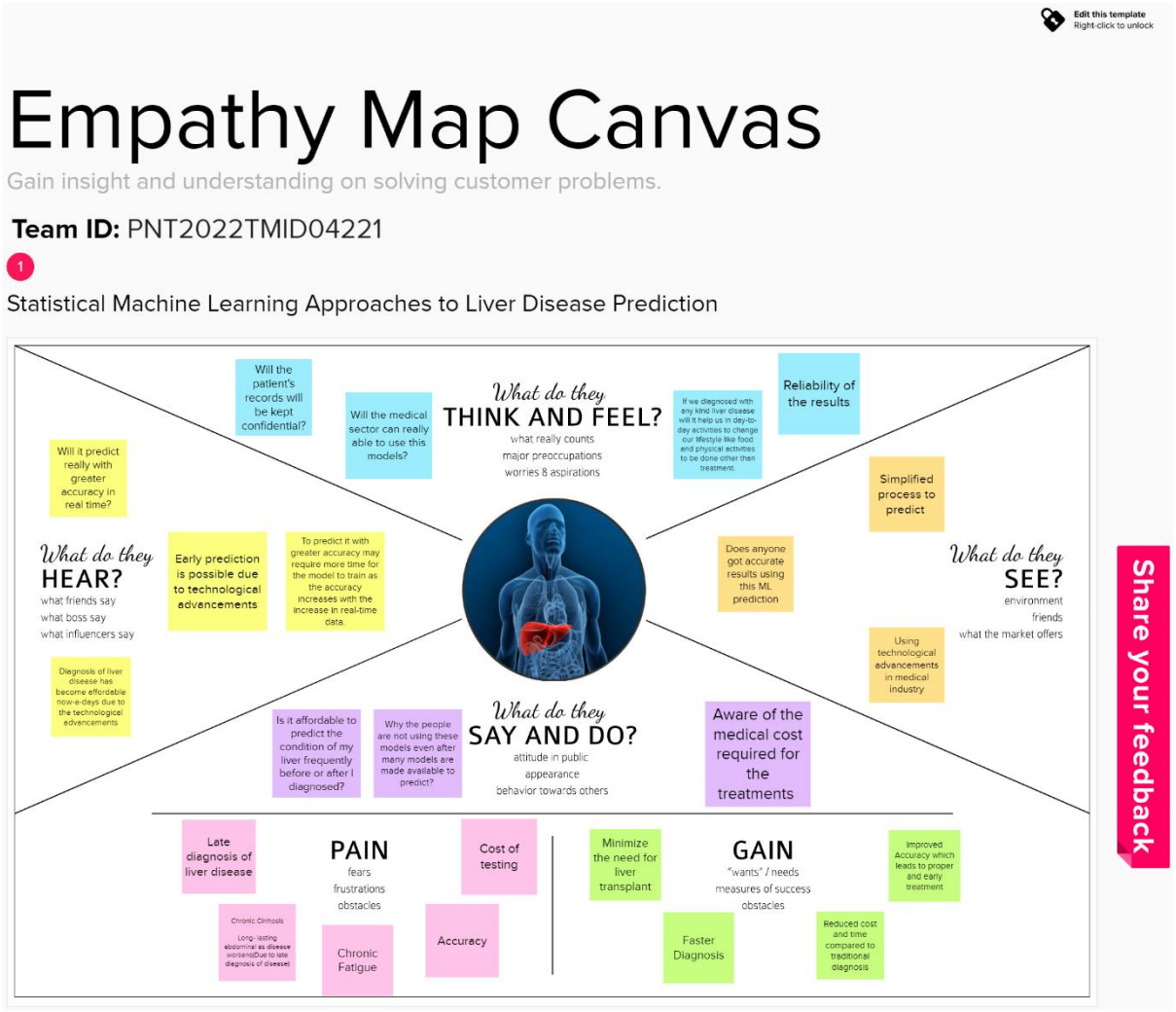
2.3 Problem Statement Definition:

Statistical Machine Learning Approaches To Liver Disease Prediction

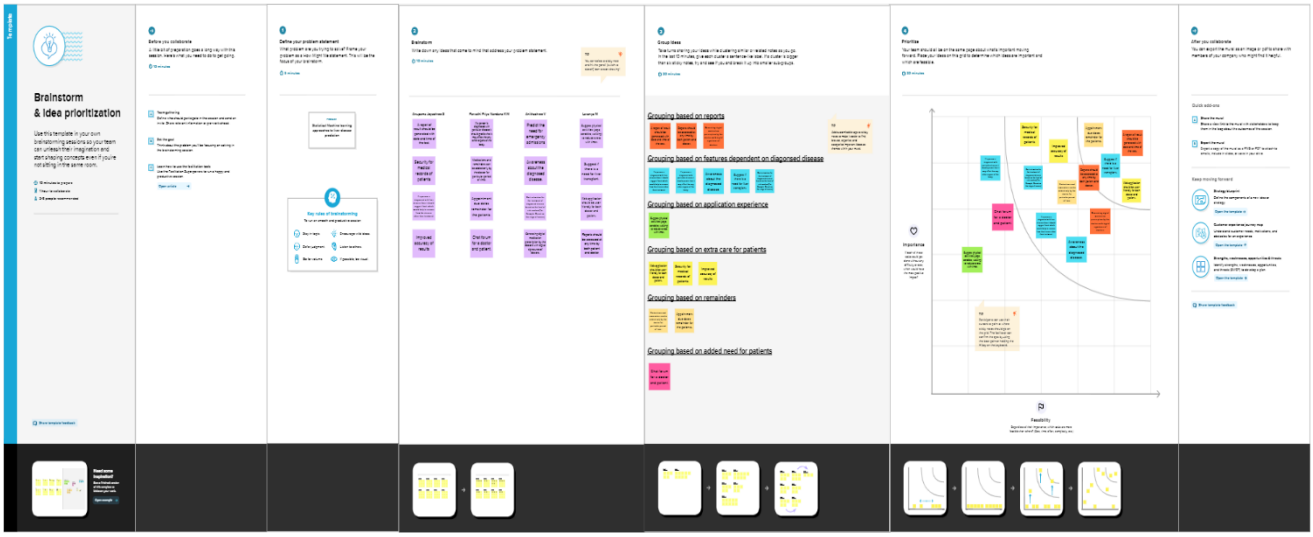
Liver Disease prevent the liver's normal function. An effective task that can assist clinicians in quickly diagnosing the disease is the early prediction of liver disease using classification algorithms. This project's main goal is to compare the predicted accuracy of different categorization algorithms by analysing their parameters. A core list of liver enzymes, proteins, age, and gender are the focus of this project's analysis of patient data from the liver, which aims to forecast the likelihood of liver illness.

3. IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation and Brainstorming:

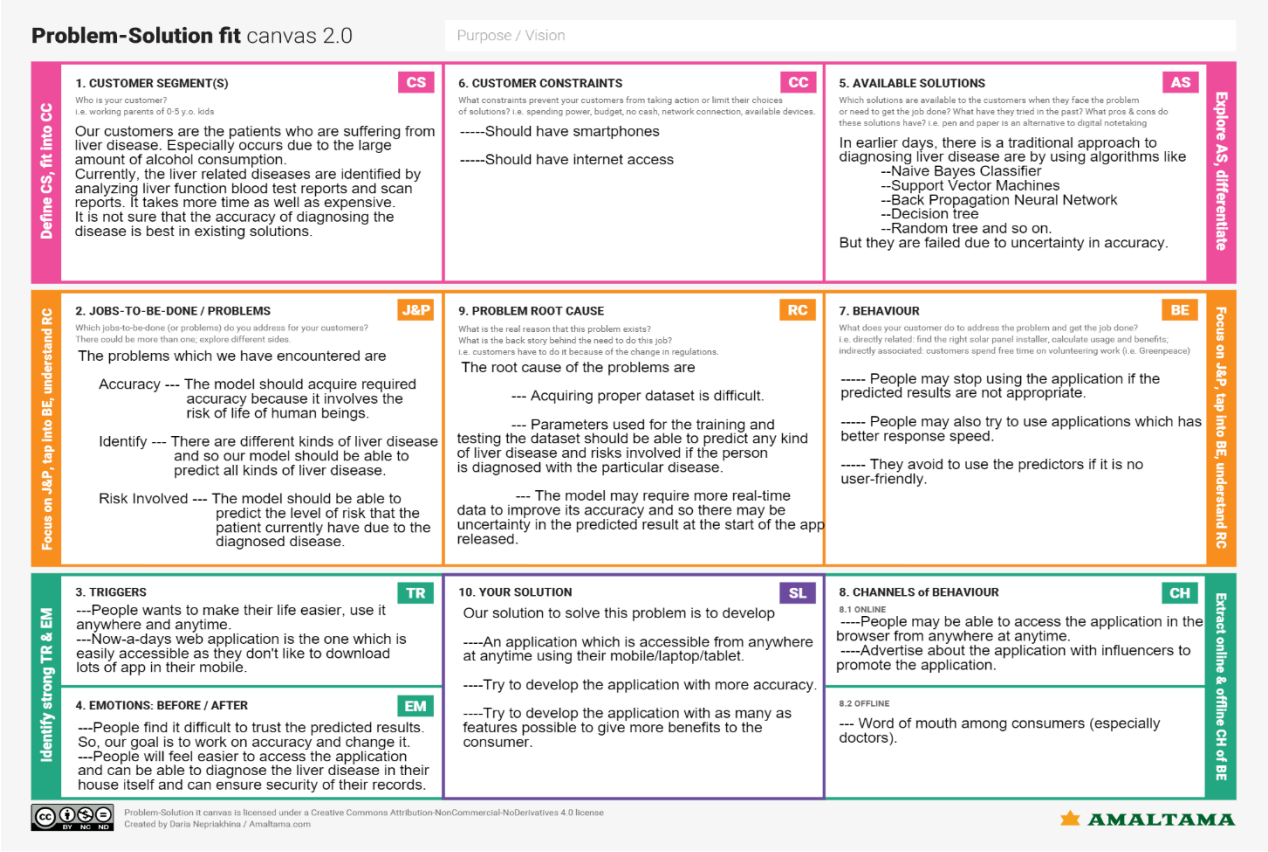


3.3 Proposed Solution

S. No	Parameter	Description
1.	Problem statement	Around 2 million people die from liver diseases each year in the world, including 1 million from cirrhosis complications, 1 million from viral hepatitis, and 1 million from hepatocellular carcinoma. A liver performs many metabolic functions, including nutrient processing and distribution. Every one in ten people get affected by liver diseases due to the adverse effects of drugs or autoimmune reactions. Common liver infections are cirrhosis, liver cancer, Wilson disease, hepatitis A and B. Acute hepatitis continues to belargely caused by drug-induced liver injury, while viralhepatitis remains highly prevalent.
2.	Idea description	Due to this, it is necessary to deal with this issue in advance. In order to avoid the loss of life due to liver diseases, early detection is crucial. Technologicaladvances have led to many applications that predict liver diseases in different technical fields. However, there are no applications that provide the most accurateresults. Therefore, the goal of the project is to identify liver disease in patients by taking list of age, proteins, and enzymes into consideration. The main model will be developed using various statistical algorithms and willbe attached with web application to provide accurate results to the patients. Using certain datasets, the modelwill be trained with algorithms, and it will be processedto predict the type of liver disease.
3.	Novelty / uniqueness	The objective of this project is to determine the best algorithm for determining liver disease at an earlystage by considering the values of different machine learning algorithms and analysing their results. Considering those machine learning algorithms like Naïve Bayes, K-Nearest Neighbour (KNN) etc., it is also necessary to consider the input values entered by the user. By analysing both the accuracy results for themodel and the machine learning algorithms, we can determine which algorithm is most effective inpredicting liver disease.
4.	Social impact / customer satisfaction	Liver disease is one of the key causes of high numbersof deaths in the country and is considered a life- threatening disease, not just anywhere, but worldwide.Liver disease can also impact people early in their life.The application will have the option to predict liver infection before and advise the wellbeing condition. These methods can reduce many of the limitations that occur in healthcare associated with inaccuracy in diagnoses, missing data, cost, and time. The application of the ML methods can help reduce the total burden of liver disease on public health worldwide by improving the recognition of risk factorsand diagnostic variables. More importantly, for chronicliver disease,

		detecting liver disease at earlier stages orin hidden cases by ML could decrease liver-related mortality, transplants, and/or hospitalizations. Earlydetection improves prognosis since treatment can be given before the progression of the disease to later stages. Invasive tests, such as biopsy, would occur lessin this case as well. Although this study focused on hepatitis and chronic liver disease variables for ML training, it can be hypothesized that the methods can beused to distinguish other types of liver disease from healthy individuals.
5.	Business model (financial benefit)	The liver is one of the most important organs in the human body. liver diseases should be found at the earlystage. So, this method is useful to detect liver disease easily in the early stage. This application is useful for both doctors and patients for the following reasons: 1. The performance classification of liver-based diseases is further improved. 2. Time complexity and accuracy can be measuredby various machine learning models, so that wecan measure different parameters, owing to theneeds of the user. 3. The cost of traditional testing is expensive and people cannot easily afford those tests. 4. Our methods can save time and costs for the betterment of people. 5. No medical expertise is required You don't needto have any knowledge of medical science. **6.** The application will have the option to predict liver infection before and advise the wellbeing condition. This application can be surprisingly gainful in low-salary nations where an absence of medicinal foundations and just particular specialists.
6.	Scalability of solution	In this project, machine learning algorithms will be used to build a model and then compared to find the most accurate models. The best accurate model will be integrated into a flask-based web application. Thus, the web application allows users to predict diseases easily by entering data they are familiar with. Upon completing the data entry, the user is instantly shown the results in their User Interface.

3.4 Problem Solution Fit



4.REQUIREMENT ANALYSIS

4.1 Functional Requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR -1	User Registration	The user registration process includes the creation of account through either email id or phone number with new password through the website.
FR - 2	User Login	The existing user can directly login to the website by giving the Login credentials.

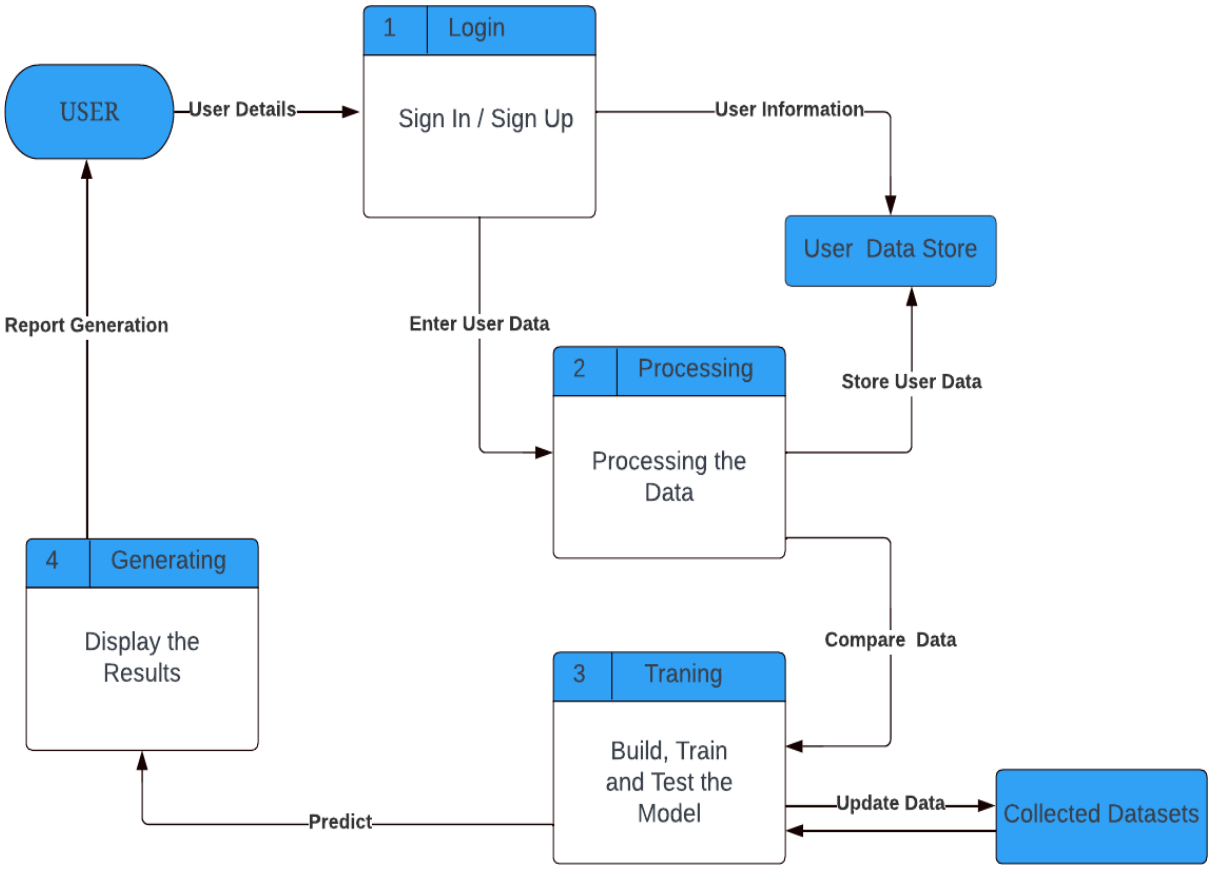
FR - 3	Admin Login	The admin can login to the website where the admin can find the analysis of the predicted data.
FR - 4	Upload Image	The user can upload the scanned image of liver in the dropdown menu from various assets like (drop box, gallery etc.,)
FR-5	Enter Data	The user can enter the required data from the scanned reports.
FR - 6	Prediction	The data entered by the user is tested with the trained model in the Watson studio to predict whether the user has liver disease or not.
FR-7	Display	The result will be displayed in theapplication.
FR-8	Report generation	The result of the tested data will be generated as report in the form of PDF andstored in the user login and also it will automatically download to the user system.

4.2 Non-Functional Requirement

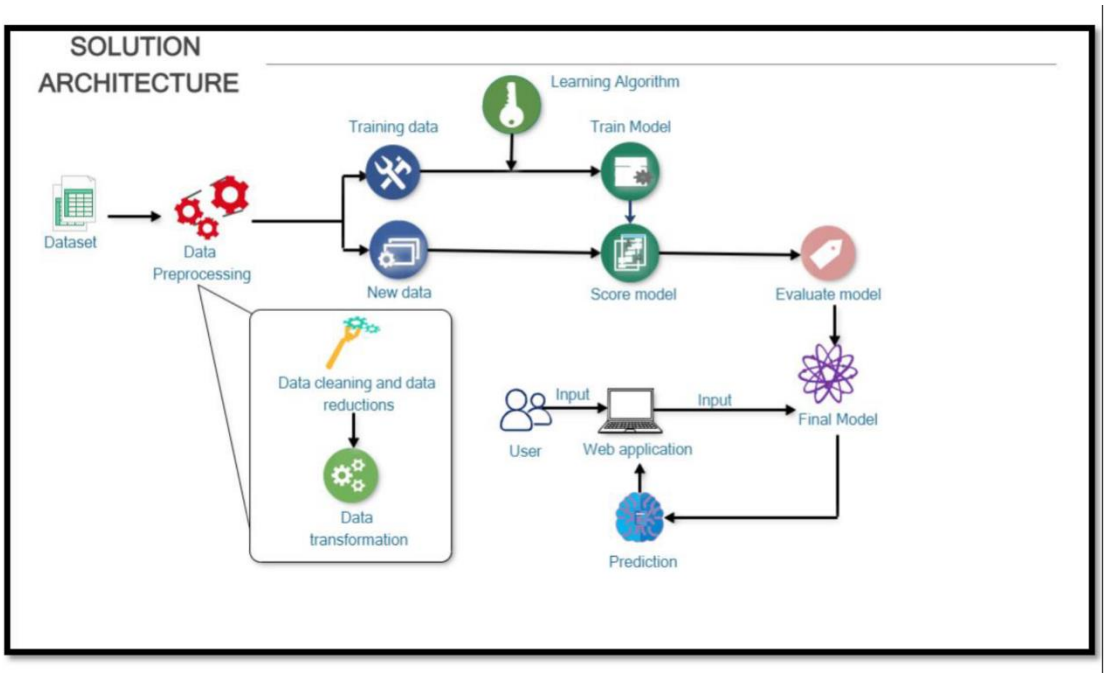
NFR No.	Non-Functional Requirement	Description
NFR - 1	**Usability**	The application can be easily accessible by any type of individuals, the aged individual and affected by liver disease can also use this tool for Diagnosis.
NFR - 2	**Security**	Data security is important to store the customer data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

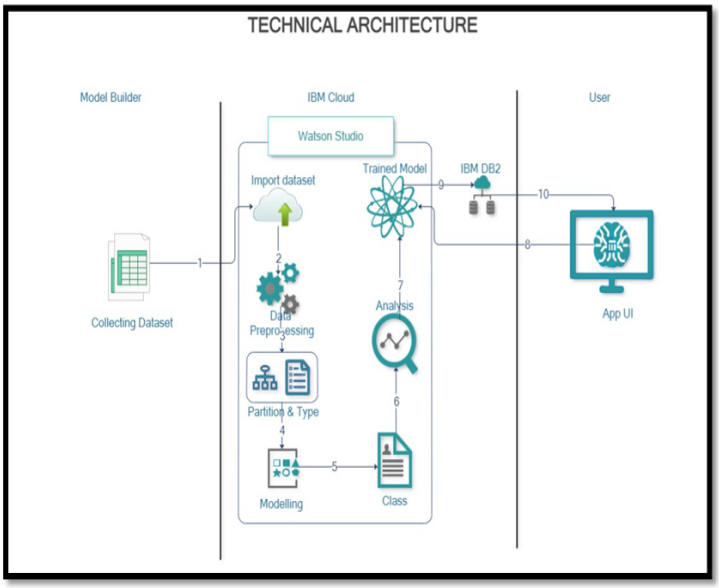
5.PROJECT DESIGN

5.1Data Flow Diagrams



5.2 Solution and Technical Architecture





1. Import the collected dataset into the Watson studio.
2. Process the datasets as per the requirements.
3. Data pre-processing can be done by partitioning and Type
   Partitioning—splits data into separate subsets for training, testing and validation stages of model building
   Type—Specify field metadata and properties that are invaluable for modelling.
4. Build the model by splitting the sample based on the field that gives maximum information gain.
5. A class of data is derived from the dataset with maximum information gain.
6. Evaluate the ability of a model to generate accurate predictions.
7. Final model is developed based on evaluation.
8. User giving their data into the app.
9. Predict the results and store it into IBM DB2 cloud database.
10. Display the results to the user in the application.

Table-1 : Components & Technologies:

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

Table-2: Application Characteristics:

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

5.3 User Stories

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

	Dashboard	USN-5	As a user, I can navigate through different pages using the dashboard	I can view various pages	High	Sprint-1
Customer (Web user)	Upload Images	USN-6	As a user, I can upload the image that required for finding whether liver disease is there are not.	Can get result based on the information provided.	Medium	Sprint-3
	Enter data	USN-7	As a user, I can enter the required data from the scanned report.	Can get result based on the information provided.	High	Sprint-3
	Report	USN-8	As a user, I can generate the report in PDF format.	Result can be generated in PDF format in user login.	Low	Sprint-4
	Search	USN-9	As a user, I can search for the specialist and best hospital in that respective field.	I can receive information on various doctors and hospitals.	Low	Sprint-4
Administrator	Analyse	USN-10	As an admin, I will analyse the given data.	I can analyse the given data.	High	Sprint-2

6.PROJECT PLANNING AND SCHEDULING

6.1 Sprint Planning and Estimation

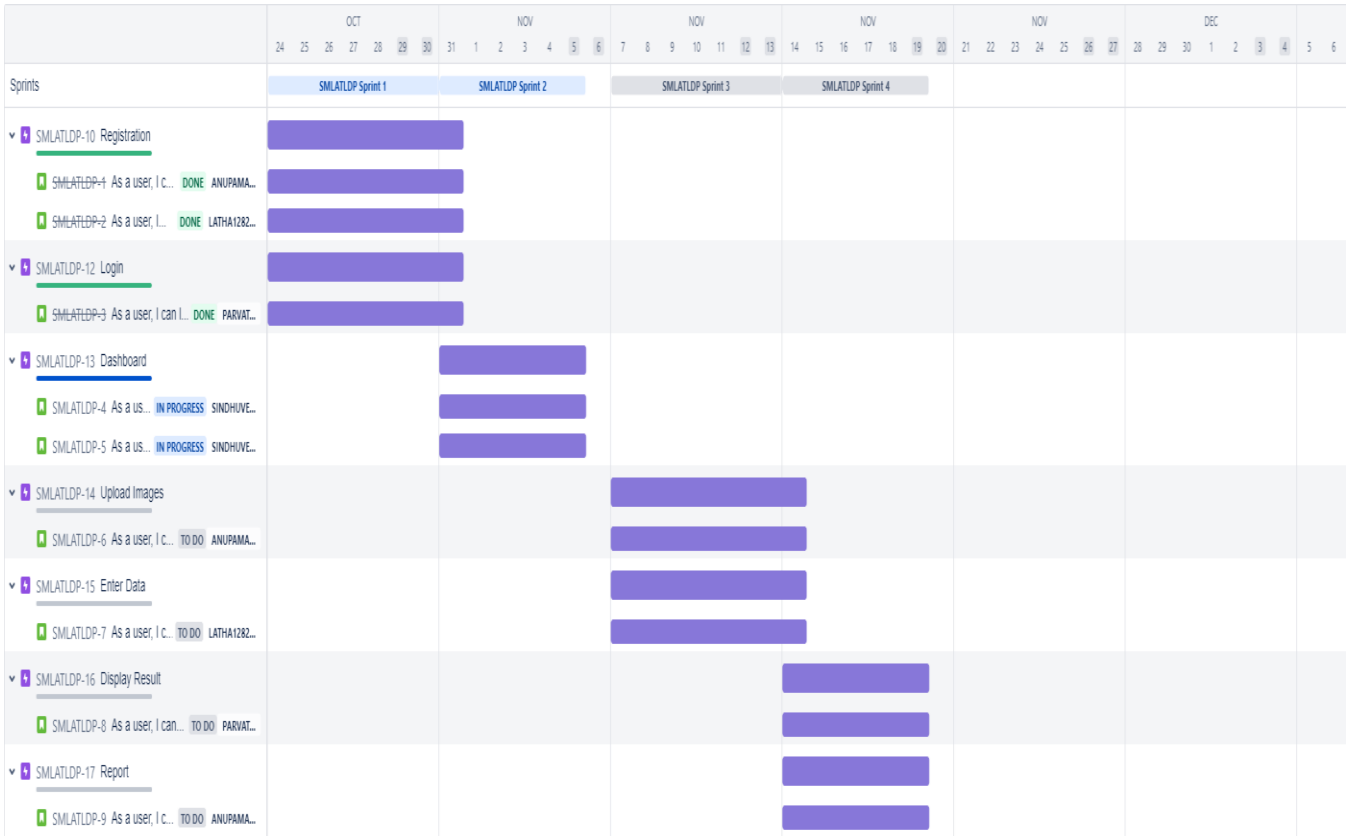
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	10	High	Anupama Jeyashree S
Sprint-1		USN-2	As a user, I will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

Sprint -2	Dashboard	USN-4	As a user, I must enter my details.	15	Medium	Anithashree V
Sprint -2	Dashboard	USN-5	As a user, I can navigate through different pages using the dashboard.	5	High	Anithashree V
Sprint - 3	Upload Images	USN-6	As a user, I can upload the image that required for finding whether liver disease is there are not.	5	Medium	Lavanya M
Sprint-3	Enter Data	USN-7	As a user, I can enter the required data from the scanned report.	15	High	Anupama Jeyashree S
Sprint - 4	Display Result	USN-8	As a user, I can view the result.	15	High	Parvathi Priya Nandana K M

6.2 Sprint Delivery Schedule

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

6.3 Reports from JIRA


OCT
NOV
NOV
NOV
NOV
DEC
24 25 26 27 28 29 30 31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 1 2 3 4 5 6
Sprints
SMLATLDP Sprint 1
SMLATLDP Sprint 2
SMLATLDP Sprint 3
SMLATLDP Sprint 4
SMLATLDP-10 Registration
SMLATLDP-1 As a user, I c... DONE ANUPAMA...
SMLATLDP-2 As a user, I... DONE LATHA1282...
SMLATLDP-12 Login
SMLATLDP-3 As a user, I can l... DONE PARVAT...
SMLATLDP-13 Dashboard
SMLATLDP-4 As a us... IN PROGRESS SINDHUVE...
SMLATLDP-5 As a us... IN PROGRESS SINDHUVE...
SMLATLDP-14 Upload Images
SMLATLDP-6 As a user, I c... TO DO ANUPAMA...
SMLATLDP-15 Enter Data
SMLATLDP-7 As a user, I c... TO DO LATHA1282...
SMLATLDP-16 Display Result
SMLATLDP-8 As a user, I can... TO DO PARVAT...
SMLATLDP-17 Report
SMLATLDP-9 As a user, I c... TO DO ANUPAMA...