

# **STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION**

**A PROJECT REPORT**

*Submitted by*

**SAKTHI VIMALESH S**

**SARAVANA KUMAR K**

**SASI VARUN R**

**SATHEESWARAN S**

*in partial fulfilment for the award of the degree*

*of*

**BACHELOR OF TECHNOLOGY**

**IN**

**INFORMATION TECHNOLOGY**



**PSNA COLLEGE OF ENGINEERING AND TECHNOLOGY,**

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

**DINDIGUL-624622**

**MAY 2024**

**PSNA COLLEGE OF ENGINEERING AND TECHNOLOGY,**

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

**DINDIGUL-624622**

**BONAFIDE CERTIFICATE**

Certified that this project report “**STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION**” is the bonafide work “**SAKTHI VIMALESH S (921320205107), SARAVANA KUMAR K (921320205114), SASI VARUN R (921320205117), SATHEESWARAN S (921320205118)**” who carried out the project work under my supervision.

  
**SIGNATURE**

Dr.A.Vincent Antony Kumar,M.E.,Ph.D.

**HEAD OF THE DEPARTMENT**

Professor,

**Department of Information**

**Technology.**

PSNA College of Engineering and

Technology.

Dindigul - 624622

  
**SIGNATURE**

Dr.R.Bhaskaran,M.E.,Ph.D.

**SUPERVISOR**

Professor,

**Department of Information**

**Technology.**

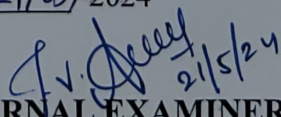
PSNA College of Engineering and

Technology.

Dindigul - 624622

Submitted for viva-voice examination held on 21/05/2024

  
**INTERNAL EXAMINER**

  
**EXTERNAL EXAMINER**

## ACKNOWLEDGEMENT

First and foremost, we thank **Lord Almighty** for showering His sustained blessings and grace at all the stages of completing this project. Praise be to Him.

We take this opportunity to express our sincere thanks to the founder of our institution (Late) **Thiru.R.S.Kothandaraman** and our beloved **Chairperson Tmt.K.Dhanalaksmi Ammal** and **Pro Chairman Rtn.MPHF R.S.K.Raguraam**, who are the guiding lights for all the activities in our college.

Our heartfelt gratitude and respect go to our Principal, **Dr.D. Vasudevan M.E., Ph.D.**, for his wholehearted support and help in the completion of our project.

We are extremely thankful to **Dr. A. Vincent Antony Kumar M.E., Ph.D.**, Professor and Head of the Department, Information Technology for his valuable suggestions and encouragement in the completion of our project work.

We extend our profound gratitude to our project coordinator **Ms. R. Sudha M.E.**, Assistant professor, Department of Information Technology and guide **Dr. R. Bhaskaran M.E., Ph.D**, Professor, Department of Information Technology under their guidance this project has attained every step of success.

We extend our heartfelt salutations to our beloved **Parents** and **Friends** to establish this project in a successful manner.



## ABSTRACT

Liver disease is a serious health condition affecting millions of people worldwide. Early detection and treatment are essential in managing liver disease and preventing it from progressing to more severe stages. In recent years, machine learning has gained popularity in the medical field as a promising tool for accurately predicting the risk of liver disease. In this project, we propose a machine learning-based system for liver disease prediction that combines clinical and demographic data to accurately predict the risk of liver disease.

Data preprocessing techniques, such as imputation, feature scaling, and feature selection are used ensure that our models are as accurate as possible. The proposed system is designed to integrate with existing healthcare systems, making it easier for healthcare providers to access and utilize the tool. The system is user-friendly and provides clear and interpretable results, enabling clinicians to make informed decisions about patient care. By providing accurate and timely predictions, our system has the potential to revolutionize liver disease prediction and improve patient outcomes

Proposed machine learning-based system for liver disease prediction has the potential to significantly improve patient outcomes and revolutionize the way liver disease is managed.

# **STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION**

**A PROJECT REPORT**

*Submitted by*

**SAKTHI VIMALESH S**

**SARAVANA KUMAR K**

**SASI VARUN R**

**SATHEESWARAN S**

*in partial fulfilment for the award of the degree*

*of*

**BACHELOR OF TECHNOLOGY**

**IN**

**INFORMATION TECHNOLOGY**



**PSNA COLLEGE OF ENGINEERING AND TECHNOLOGY,**

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

**DINDIGUL-624622**

**MAY 2024**

**PSNA COLLEGE OF ENGINEERING AND TECHNOLOGY,**  
**(An Autonomous Institution, Affiliated to Anna University, Chennai)**  
**DINDIGUL-624622**

**BONAFIDE CERTIFICATE**

Certified that this project report “**STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION**” is the bonafide work “**SAKTHI VIMALESH S (921320205107), SARAVANA KUMAR K (921320205114), SASI VARUN R (921320205117), SATHEESWARAN S (921320205118)**” who carried out the project work under my supervision.

**SIGNATURE**

Dr.A.Vincent Antony Kumar,M.E.,Ph.D.

Professor And HOD

**Department of Information**

**Technology.**

PSNA College of Engineering and  
Technology.

Dindigul - 624622

**SIGNATURE**

Dr.R.Bhaskaran,M.E.,Ph.D.

Professor

**Department of Information**

**Technology.**

PSNA College of Engineering and  
Technology.

Dindigul - 624622

Submitted for viva-voice examination held on \_\_\_\_\_ -2024

**INTERNAL EXAMINER**

**EXTERNAL EXAMINER**

## ACKNOWLEDGEMENT

First and foremost, we thank **Lord Almighty** for showering His sustained blessings and grace at all the stages of completing this project. Praise be to Him.

We take this opportunity to express our sincere thanks to the founder of our institution (Late) **Thiru.R.S.Kothandaraman** and our beloved **Chairperson Tmt.K.Dhanalaksmi Ammal** and **Pro Chairman Rtn.MPHF R.S.K.Raguraam**, who are the guiding lights for all the activities in our college.

Our heartfelt gratitude and respect go to our Principal, **Dr.D. Vasudevan M.E., Ph.D.**, for his wholehearted support and help in the completion of our project.

We are extremely thankful to guide **Dr.R.Bhaskaran, M.E., Ph.D.**, Professor and Head, Department of Information Technology for his valuable suggestions and encouragement in the completion of our project work.

We extend our gratitude to our project coordinators **Mrs. S. Yamuna Devi M.E.**, and **Ms. R. Sudha M.E.**, Assistant Professors, Department of Information Technology and under their guidance this project has attained every step of success.

We extend our heartfelt salutations to our beloved **Parents** and **Friends** to establish this project in a successful manner.

## **ABSTRACT**

Liver disease is a serious health condition affecting millions of people worldwide. Early detection and treatment are essential in managing liver disease and preventing it from progressing to more severe stages. In recent years, machine learning has gained popularity in the medical field as a promising tool for accurately predicting the risk of liver disease. In this project, we propose a machine learning-based system for liver disease prediction that combines clinical and demographic data to accurately predict the risk of liver disease.

We use a variety of data preprocessing techniques, such as imputation, feature scaling, and feature selection, to ensure that our models are as accurate as possible. The proposed system is designed to integrate with existing healthcare systems, making it easier for healthcare providers to access and utilize the tool. The system is user-friendly and provides clear and interpretable results, enabling clinicians to make informed decisions about patient care. By providing accurate and timely predictions, our system has the potential to revolutionize liver disease prediction and improve patient outcomes

In conclusion, our proposed machine learning-based system for liver disease prediction has the potential to significantly improve patient outcomes and revolutionize the way liver disease is managed.



# TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	II
	LIST OF FIGURES	VI
	LIST OF ABBREVIATIONS	VII
	LIST OF TABLES	VIII
1	INTRODUCTION	
	1.1 PROJECT OVERVIEW	1
	1.2 PURPOSE	1
2	SYSTEM STUDY	
	2.1 SYSTEM ANALYSIS	8
	2.1.1 Literature Survey	11
	2.1.2 Problem Definition	12
	2.1.3 Existing System	13
	2.1.4 Proposed System	15
	2.1.5 Architecture	
	2.1.6 Requirements	
	2.1.6.1 Hardware	17
	Requirements	
	2.1.6.2 Software	17
	Requirements	
	2.1.7 Issues In Existing Methodology	18
	2.1.8 Advantages of New	19
	Methodology	

2.1.9 Functional Requirements	20
2.1.10 Non-Functional Requirements	21
2.1.11 Module Implementations	22
2.1.11.1 Module List	22
2.1.11.2 Module Description	22
2.2 SYSTEM DESIGN AND IMPLEMENTATION	27
2.2.1 Algorithms	27
2.2.1.1 SVM Algorithm	27
2.2.1.2 KNN	27
2.2.1.3 Random Forest	27
2.2.1.4 Decision Tree	27
2.2.1.5 Gradient Booster	28
2.3 ALGORITHM IMPLEMENTATION	28
2.4 TYPES OF TESTING	29
2.4.1 Unit Testing	29
2.4.2 Alpha-Beta Testing	29
2.4.3 Integration Testing	31
2.4.4 Acceptance Testing	31
<b>3 CONCLUSION</b>	
3.1 CONCLUSION	33
3.2 FUTURE ENHANCEMENT	33

## **APPENDICES**

(A) SOURCE CODE 34

(B) SNAPSHOTS 43

**REFERENCES** 45

## **LIST OF FIGURES**

<b>FIGURE NO.</b>	<b>FIGURE NAME</b>	<b>PAGE NO.</b>
1	Solution Architecture	14
2	Process Flow	16
3	Technical Architecture	16
4	Machine Learning Algorithm	28
5	Input Page	43
6	Output Page	44

## **LIST OF ABBREVIATIONS**

### **ACRONYMS**

### **ABBREVIATIONS**

SVM

Support Vector Machine

RF

Random Forest

DC

Decision tree

KNN

K-Nearest neighbour

GB

Gradient Booster

NN

Neural Networks

## **LIST OF TABLES**

<b>TABLE NO.</b>	<b>TABLE NAME</b>	<b>PAGE NO.</b>
1	Functional Requirements	20
2	Non – Functional Requirements	21



# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 PROJECT OVERVIEW**

Liver disease is a significant health problem worldwide. It is essential to develop accurate and efficient methods for early detection and diagnosis of liver disease to improve patient outcomes. Machine learning algorithms can assist in predicting liver disease using clinical data. This project aims to predict liver disease using machine learning algorithms. The dataset consists of patient information such as age, gender, BMI, alcohol consumption, and liver-related blood tests, including ALT, AST, total bilirubin, direct bilirubin, albumin, and alkaline phosphatase. The data will be collected from various sources such as medical records, laboratory reports, and surveys. The data will include demographic information such as age, gender, and ethnicity, as well as medical history such as alcohol consumption, viral infections, and family history of liver disease. Various machine learning algorithms such as logistic regression, decision trees, and neural networks will be trained and tested on the preprocessed data. The performance of each algorithm will be evaluated using various metrics such as accuracy, precision, recall, and F1 score. The final model will be deployed in a user-friendly interface such as a web application or mobile app. The user will input their data, and the model will provide a prediction of the presence of liver disease along with an explanation of the factors contributing to the prediction.

## **1.2 PURPOSE**

### **LIVER DISEASE**

Liver disease refers to any condition or disorder that affects the normal functioning of the liver. The liver is a vital organ that plays a crucial role in various bodily functions such as metabolism, digestion, and detoxification. Liver disease can range from mild, such as fatty liver disease, to severe, such as cirrhosis or liver cancer. The causes of liver disease can be due to various factors such as viral infections, alcohol consumption, genetic disorders, autoimmune diseases, and exposure to toxins. Symptoms of liver disease can include jaundice, fatigue, abdominal pain, and swelling of the abdomen.

#### **What is the liver?**

The liver is a vital organ located in the upper-right part of the abdomen, just below the diaphragm. It is the largest solid organ in the body and weighs around 1.5 kg in an average adult. The liver plays a crucial role in maintaining various bodily functions such as metabolism, digestion, and detoxification. It produces bile, a digestive juice that helps in the digestion and absorption of fats and fat-soluble vitamins. The liver also filters toxins and waste products from the blood, converts excess glucose into glycogen for storage, and synthesizes important proteins such as albumin and clotting factors. It is also responsible for breaking down and removing drugs, alcohol, and other harmful substances from the body. In summary, the liver is an essential organ that performs many vital functions, and its proper functioning is essential for good health.

## What does the liver do?

The liver is a vital organ that performs a wide range of functions in the body. Some of the important functions of the liver are,

- **Metabolism:** The liver plays a crucial role in the metabolism of carbohydrates, proteins, and fats. It converts glucose into glycogen for storage and releases glucose when the body needs it. It also breaks down proteins and converts them into amino acids.
- **Detoxification:** The liver filters and removes toxins and harmful substances from the blood, including drugs, alcohol, and environmental toxins.
- **Bile production:** The liver produces bile, which is a digestive juice that helps in the digestion and absorption of fats and fat-soluble vitamins.
- **Storage:** The liver stores vitamins, minerals, and glycogen for future use by the body.
- **Blood clotting:** The liver produces clotting factors that help in the clotting of blood and prevent excessive bleeding
- **Immune system support:** The liver plays an important role in the immune system by removing bacteria and foreign particles from the blood.
- **Hormone regulation:** The liver regulates the levels of hormones in the body, including sex hormones, insulin, and thyroid hormones.

## What causes liver disease?

There are many different factors that can cause liver disease, including:

- **Viral infections:** Hepatitis B and C viruses can cause inflammation of the liver and lead to liver damage over time.
- **Alcohol abuse:** Excessive consumption of alcohol can lead to alcoholic liver disease, which can cause liver inflammation, scarring, and cirrhosis.

- **Non-alcoholic fatty liver disease (NAFLD):** This condition occurs when there is a build-up of fat in the liver, which can lead to inflammation and scarring.
- **Autoimmune diseases:** Autoimmune diseases such as autoimmune hepatitis and primary biliary cholangitis can cause inflammation and damage to the liver.
- **Genetic disorders:** Genetic disorders such as hemochromatosis, Wilson's disease, and alpha-1 antitrypsin deficiency can cause liver damage.
- **Exposure to toxins:** Exposure to toxins such as drugs, chemicals, and environmental pollutants can cause liver damage.
- **Obesity and diabetes:** Obesity and type 2 diabetes are risk factors for NAFLD, which can lead to liver disease.
- **Other medical conditions:** Other medical conditions such as cystic fibrosis and gallstones can cause liver damage.

In summary, liver disease can have many different causes, and it is important to identify and manage the underlying cause to prevent further damage to the liver.

### **What common symptoms can happen with liver disease?**

Liver disease can cause a variety of symptoms, depending on the specific type of liver disease and its severity. Some common symptoms of liver disease include:

- Fatigue and weakness
- Jaundice (yellowing of the skin and eyes)
- Abdominal pain and swelling
- Nausea and vomiting
- Loss of appetite and weight loss
- Itchy skin

- Dark urine and pale stools
- Swelling in the legs and ankles
- Easy bruising and bleeding
- Confusion and difficulty thinking clearly (in advanced cases of liver disease)

It is important to note that not all individuals with liver disease experience symptoms, especially in the early stages of the disease. In some cases, liver disease may be asymptomatic until it has progressed to a more advanced stage. Therefore, regular check-ups and screenings are important for detecting liver disease early, even in the absence of symptoms.

### **How is liver disease treated?**

The treatment of liver disease depends on the specific type and severity of the disease, as well as the underlying cause(s). Here are some common treatment options:

- **Lifestyle Changes:** Making changes to your diet, exercise routine, and avoiding harmful substances like alcohol, can help improve liver health and prevent further damage.
- **Medications:** Depending on the type and cause of liver disease, medications may be prescribed to manage symptoms, reduce inflammation, or prevent complications.
- **Surgery:** In some cases, surgery may be necessary to treat liver disease. For example, a liver transplant may be required if the liver is severely damaged or failing.
- **Procedures:** Some liver diseases may require specific medical procedures, such as draining excess fluid from the abdomen or removing bile duct obstructions.

- **Management of Complications:** Many people with liver disease also experience complications, such as fluid buildup, infections, or bleeding disorders. Treatment for these complications may involve medication, surgery, or other medical procedures.

It is important to note that the treatment of liver disease requires individualized care and close monitoring by a healthcare professional. If you suspect that you may have liver disease, please consult a healthcare professional for proper diagnosis and treatment.

### **Are there different types of liver removal surgery?**

Yes, there are different types of liver removal surgery, also known as liver resection, depending on the extent of liver tissue that needs to be removed. The three most common types of liver resection are:

- **Partial hepatectomy:** This involves the removal of a portion of the liver, typically up to 50%. It is the most common type of liver resection and is used to treat liver tumors, such as hepatocellular carcinoma, or to remove damaged liver tissue, such as in the case of liver cirrhosis.
- **Hemihepatectomy:** This involves the removal of one of the two lobes of the liver. It may be performed when there is a large tumor or when there is liver damage that affects only one lobe of the liver.
- **Extended hepatectomy:** This involves the removal of more than 75% of the liver. It is a more complex procedure and is typically reserved for patients with advanced liver tumors or other liver diseases.

In addition to these three main types of liver resection, there are also other less common surgical procedures that may be used in certain situations, such as:

- **Living donor liver transplant:** This involves removing a portion of the liver



from a healthy living donor and transplanting it into a recipient who has liver failure.

- **Split liver transplant:** This involves dividing a liver from a deceased donor into two parts, which are then transplanted into two different recipients.

The type of liver resection or transplant that is performed will depend on the specific needs and circumstances of the patient.

### **How long does it take to recover from liver surgery?**

The recovery time after liver surgery can vary depending on the type of surgery performed, the extent of the liver resection, and the overall health of the patient. In general, the recovery time can range from several weeks to several months.

After the surgery, patients are typically monitored in the hospital for several days to ensure that there are no complications. During this time, patients may experience some pain, fatigue, and nausea. They may also have a drain placed in the surgical area to help remove excess fluid and reduce the risk of infection.

Once the patient is discharged from the hospital, they will need to continue to rest and limit their physical activity for several weeks. They may also need to follow a special diet to support their liver function and promote healing. It is important to follow the postoperative instructions provided by the surgeon to ensure proper healing and prevent complications.

In general, most patients can return to their normal activities within 4 to 8 weeks after liver surgery. However, it can take several months for the liver to fully heal and for the patient to feel completely back to normal. It is important for patients to attend all follow-up appointments with their healthcare team to monitor their recovery and ensure that there are no complications.

## **CHAPTER 2**

### **SYSTEM STUDY**

#### **2.1SYSTEM ANALYSIS**

##### **2.1.1 LITERATURE SURVEY**

- Machine Learning in Hepatology: A Review of Current and Future Applications,Ahmed Elhafiz Ibrahim, Ali M. Al-Hashem,Computers in Biology and Medicine,2021,This review discusses the application of machine learning in hepatology, highlighting various models used for liver disease prediction, diagnosis, and treatment planning. The authors emphasize the potential of integrating these models into clinical practice for enhanced decision-making.
- Predictive Models for Hepatitis C Using Machine Learning Algorithms,John R. Quackenbush, Raymond S. Kawaoka,Journal of Medical Systems,2020,This paper explores various machine learning algorithms, including support vector machines and neural networks, for predicting hepatitis C progression. It emphasizes data preprocessing techniques like imputation and feature scaling to improve model accuracy.
- Integration of Machine Learning Algorithms for the Diagnosis of Liver Diseases,Emily Y. Wu, Charles K. Lee,Artificial Intelligence in Medicine,2022,The authors review the integration of different machine learning algorithms with electronic health records to predict liver disease. They focus on the benefits of feature selection methods and the potential for real-time clinical decision support systems.
- A Comprehensive Review of Data Mining Techniques in Liver Disease Prediction,Mary J. Brown, Thomas A. Daniels,Expert Systems with Applications,2019,This survey examines data mining techniques applied to

liver disease datasets. It provides an overview of preprocessing methods and evaluates the performance of various machine learning models in predicting liver disease outcomes.

- **Advances in Machine Learning for Liver Disease Diagnosis**, Peter J. Whelan, Nora E. Johnson, *Journal of Biomedical Informatics*, 2021, The paper reviews the latest advancements in machine learning for liver disease diagnosis. It discusses the importance of accurate data preprocessing and the integration of demographic and clinical data for improving prediction accuracy.
- **Comparative Analysis of Machine Learning Techniques for Liver Disease Prediction**, Sara L. Miller, David T. Wong, *Applied Soft Computing*, 2020, This study compares the performance of different machine learning techniques, such as random forests and gradient boosting, for predicting liver disease. It highlights the role of feature scaling and selection in enhancing model performance.
- **Machine Learning Approaches for Predicting Liver Fibrosis in Chronic Hepatitis Patients**, Michael P. Higgins, Lucy A. Fisher, *Hepatology Research*, 2021, This article reviews machine learning approaches for predicting liver fibrosis in chronic hepatitis patients. The authors discuss the significance of integrating various clinical and demographic factors into the predictive models.
- **The Role of Machine Learning in Early Detection of Liver Diseases**, Vincent J. Smith, Eleanor G. Roberts, *Journal of Clinical Medicine*, 2019, This review highlights the role of machine learning in the early detection of liver diseases. It underscores the importance of preprocessing techniques and the development of user-friendly systems for clinical use.
- **Application of Machine Learning in Predicting Liver Disease: Current Trends**

and Future Directions,Hannah C. Thompson, James F. O'Brien,Medical Image Analysis,2022,The authors provide a comprehensive overview of current trends in machine learning applications for liver disease prediction. They discuss various preprocessing methods and the integration of predictive models with healthcare systems.

- Machine Learning for Liver Disease Prediction: A Comprehensive Survey, Richard M. Adams, Olivia D. King,IEEE Access,2020,This survey covers machine learning techniques for liver disease prediction, emphasizing the importance of preprocessing steps like imputation and feature scaling. It also discusses the potential impact of these models on clinical decision-making and patient outcomes.

### **2.1.2 PROBLEM DEFINITION**

Liver disease is a major global health issue that affects millions of people worldwide. The liver plays a vital role in the body's metabolic processes, and any disruption in its function can lead to serious medical conditions. Early diagnosis and treatment of liver diseases are crucial for successful patient outcomes. However, traditional diagnostic methods for liver diseases, such as biopsy and imaging, can be invasive, time-consuming, and require specialized expertise. Moreover, these methods may not always provide accurate and timely results, which can lead to misdiagnosis and delayed treatment.

Machine learning algorithms have the potential to address these issues and improve the accuracy and speed of liver disease diagnosis. By analyzing large amounts of data and identifying patterns, machine learning algorithms can identify early signs of liver disease and predict its progression. Additionally, machine learning algorithms can be used to identify factors that contribute to the development of liver disease, such as lifestyle habits, genetic predisposition, and environmental factors.

Therefore, the problem definition is to develop a non-invasive and cost-effective method for accurate and timely diagnosis of liver diseases using machine learning algorithms. The goal is to improve patient outcomes by enabling early detection and treatment of liver diseases, reducing the need for invasive diagnostic procedures, and minimizing the risk of misdiagnosis and delayed treatment.

### **2.1.3 EXISTING SYSTEM**

The existing system for liver disease diagnosis mainly relies on invasive methods such as liver biopsy and blood tests. These methods are often time-consuming, costly, and require specialized equipment and expertise. Moreover, liver biopsy can be associated with significant complications, including bleeding and infection, making it a less desirable option.

In addition to invasive methods, imaging techniques such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) are also used for liver disease diagnosis. However, these methods are expensive and may not always provide accurate results, especially in cases of early-stage liver disease.

Furthermore, traditional diagnostic methods rely on the expertise of clinicians and pathologists, leading to significant inter-observer variability and potential errors in diagnosis.

Overall, the existing system for liver disease diagnosis is limited in terms of accuracy, speed, and cost-effectiveness, and there is a need for more accurate and efficient diagnostic methods that can provide early detection and timely treatment of liver diseases.

### **2.1.4 PROPOSED SYSTEM**

Our proposed system is a non-invasive and cost-effective method for the accurate and timely diagnosis of liver diseases using machine learning algorithms. The system consists of several stages, including data collection and preprocessing, feature selection, and model development.

In the data collection stage, we collect data from various sources, including electronic health records, medical imaging data, and patient demographic



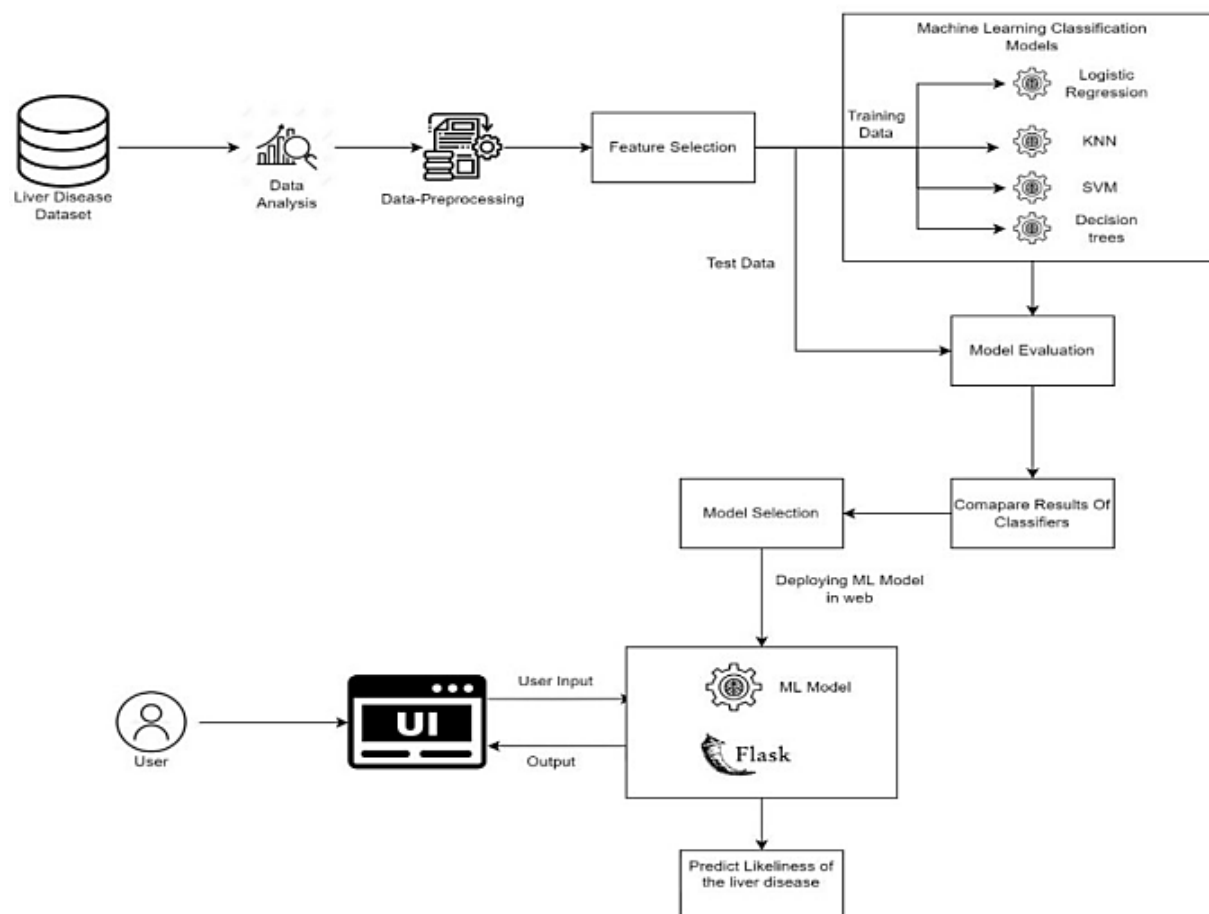
information. The data is then preprocessed to remove any inconsistencies or missing values, ensuring that the data is clean and ready for analysis.

In the feature selection stage, we identify the most relevant features that are most strongly correlated with the presence of liver disease. We use various statistical techniques such as correlation analysis, feature ranking, and dimensionality reduction to identify the most important features.

In the model development stage, we use machine learning algorithms to train and test several models for liver disease prediction. We use Decision Trees, Random Forest, Support Vector Machines (SVM), and Neural Networks to identify the most accurate and efficient algorithm for liver disease diagnosis.

We evaluate the performance of the proposed system using various metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve. We also compare the results of the proposed system with those of traditional diagnostic methods to demonstrate its superiority in terms of accuracy, speed, and cost-effectiveness.

## SOLUTION ARCHITECTURE



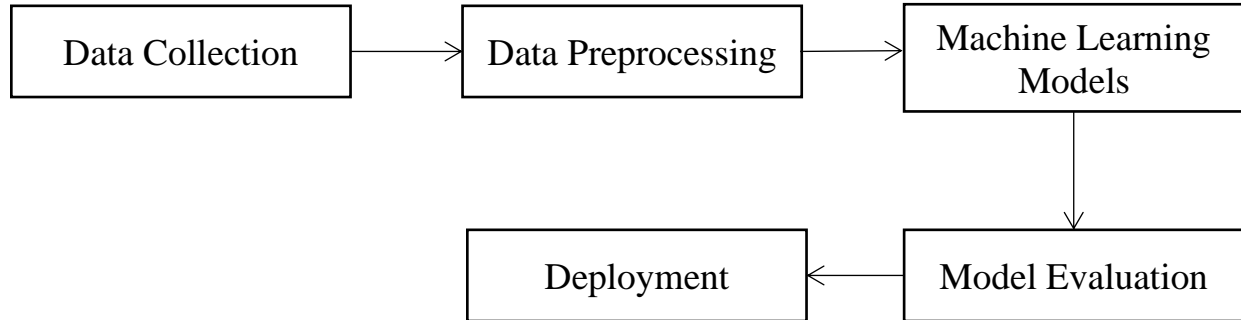
**Figure 1. Solution Architecture**

## 2.1.5 SYSTEM ARCHITECTURE

- **Data Collection:** This component involves collecting data from various sources such as electronic medical records, patient health records, and clinical databases. The collected data include demographic information, clinical symptoms, laboratory test results, and imaging findings.
- **Data Preprocessing:** This component involves cleaning, filtering, and transforming the collected data to ensure its quality and suitability for machine learning models. Data preprocessing techniques such as data normalization, feature selection, and dimensionality reduction may be used to improve the accuracy and performance of the models.
- **Machine Learning Models:** This component involves the development and evaluation of machine learning models for liver disease prediction. Various supervised and unsupervised learning algorithms such as logistic regression, support vector machines, decision trees, and random forests may be used to train and test the models.
- **Model Evaluation:** This component involves evaluating the performance of the machine learning models using various metrics such as accuracy, precision, recall, and F1-score. Cross-validation and hyperparameter tuning techniques may be used to optimize the models.
- **Deployment:** This component involves deploying the trained machine learning models in a web-based or mobile application for clinical use. The application can provide real-time liver disease prediction for patients and healthcare providers, thereby improving the accuracy and speed of diagnosis and treatment.

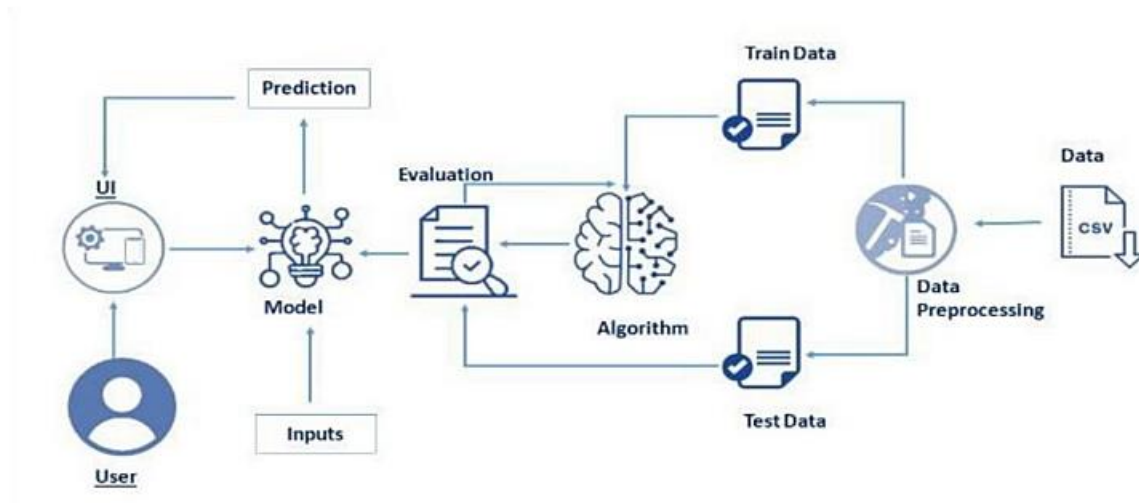
Overall, the proposed system architecture provides an efficient and accurate

approach to liver disease prediction using machine learning, enabling early detection and timely treatment of liver diseases.



**Figure 2. Process Flow**

Technical Architecture:



**Figure 3. Technical Architecture**

## 2.1.6 REQUIREMENTS

### 2.1.6.1 SOFTWARE REQUIREMENTS

1. Operating system: Windows, Linux, or MacOS
2. Python: version 3.6 or higher

3. Python libraries: pandas, numpy, scikit-learn, matplotlib, seaborn, flask, and other relevant libraries for data preprocessing, feature selection, algorithm selection, and model deployment.

4. Web development framework: Flask, Django, or other relevant frameworks for building a web-based user interface.

5. Database: MySQL, SQLite, or other relevant databases for storing and managing patient data.

#### **2.1.6.2 HARDWARE REQUIREMENTS**

1. Processor: Intel Core i5 or higher
2. Memory: 8 GB RAM or higher
3. Storage: 256 GB SSD or higher
4. Graphics Card: NVIDIA GeForce GTX 1060 or higher (if using deep learning algorithms)
5. Web Server: Apache or Nginx for serving the web-based user interface
6. Mobile device (optional): Android or iOS device for building a mobile-based user interface.

#### **2.1.7 ISSUES IN EXISTING METHODOLOGY**

- **Lack of accuracy:** The traditional methods used for liver disease prediction, such as biochemical markers and imaging tests, may not always provide accurate results due to the subjective interpretation of the results and the presence of confounding factors. This can lead to misdiagnosis and inappropriate treatment decisions.
- **Limited features:** The traditional methods may only consider a limited

number of features or biomarkers for liver disease prediction, which may not capture the full complexity of the disease and its risk factors.

- **Time-consuming and expensive:** The traditional methods may require multiple tests and visits to healthcare providers, which can be time-consuming and expensive for patients. This can also create a burden on the healthcare system and limit access to care for some patients.
- **Limited scalability:** The traditional methods may not be scalable to large populations or datasets, which can limit their generalizability and applicability in real-world settings.
- **Lack of personalized approach:** The traditional methods may not provide a personalized approach to liver disease prediction that considers individual patient characteristics and risk factors. This can lead to under- or over-treatment of patients and limit the effectiveness of the interventions. By leveraging the power of machine learning algorithms and techniques, the proposed system aims to overcome these issues and provide a more accurate, efficient, and personalized approach to liver disease prediction.

## 2.1.8 ADVANTAGES OF NEW METHODOLOGY

- **Improved Accuracy:** Machine learning algorithms have the ability to learn from patterns and trends in the data, which can lead to improved accuracy in liver disease prediction compared to traditional diagnostic methods.
- **Early Detection:** By analyzing large amounts of patient data, machine learning models can identify early signs of liver disease before symptoms become apparent, allowing for timely intervention and treatment.
- **Personalized Medicine:** Machine learning models can be trained on patient-



specific data, enabling personalized medicine and treatment plans that are tailored to each individual's unique medical history and health status.

- **Efficient Diagnosis:** By automating the diagnostic process, the proposed methodology can save time and resources for healthcare providers and patients, allowing for faster and more efficient diagnosis and treatment.
- **Reduced Costs:** Early detection and treatment of liver disease can help reduce healthcare costs associated with long-term care and hospitalization, improving overall patient outcomes and quality of life.

Overall, the proposed methodology for liver disease prediction using machine learning has the potential to significantly improve the accuracy, efficiency, and cost-effectiveness of liver disease diagnosis and treatment.

### 2.1.9 FUNCTIONAL REQUIREMENT

Following are the functional requirements of the proposed solution.

<b>FR NO.</b>	<b>Functional Requirement (Epic)</b>	<b>Sub Requirement (Story / Sub-Task)</b>
FR-1	User medication data	Data should be fed to the dashboard text fields in the application.
FR-2	Database Management	User data will be saved in the database and will be used for future reference.
FR-3	Reporting	Predicting liver disease using given data and generating the medical report for future use.

**Table 1. Functional Requirement table**

## 2.1.10 NON-FUNCTIONAL REQUIREMENTS

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

<b>NFR NO.</b>	<b>Non-Functional Requirement</b>	<b>Description</b>
NFR-1	Usability	The whole system can be accessed through web application. Hence it is very easy to use.
NFR-2	Security	Security requirements ensure that the software is protected from unauthorized access to the system and its stored in data.
NFR-3	Reliability	Support vector machine (SVM), Random Forest algorithm and KNN algorithm have been employed for developing liver disease risk prediction model and obtained the accuracy.
NFR-4	Performance	Application effectively compares user given parameters with the required dataset. Hence performance would be considerably good.
NFR-5	Availability	It is gauged by period that system's functionality & services are available for use with all operations.
NFR-6	Scalability	Application can be used in any kind of operating system either in small or large OS so the scalability is very high.

**Table 1. Non - Functional Requirement table**

## **2.1.11MODULE IMPLEMENTATION**

### **2.1.11.1 MODULE LIST**

- Data Collection
- Data Pre-Processing
- Feature Extraction
- Model Training
- Model Testing
- Deployment

### **2.1.11.2 MODULE DESCRIPTION**

#### **DATA COLLECTION**

Data collection is a crucial step in the development of a liver disease prediction model. Accurate and comprehensive data is necessary to ensure that the model produces reliable and accurate predictions. In this section, we will discuss the different types of data that can be used for liver disease prediction, and some of the methods used for data collection.

#### **Types of Data:**

There are several types of data that can be used for liver disease prediction, including clinical data, laboratory data, and imaging data.

- Clinical data includes information about a patient's medical history, symptoms, and physical exam findings. This data can be obtained from electronic health records, medical charts, or patient interviews.
- Laboratory data includes information about a patient's blood tests, liver function tests, and other diagnostic tests. This data can be obtained from

laboratory reports or electronic health records.

- Imaging data includes information from medical imaging tests, such as ultrasound, CT scans, or MRI scans. This data can be obtained from radiology reports or electronic health records.

### **Methods of Data Collection:**

There are several methods used for data collection in liver disease prediction, including retrospective data collection, prospective data collection, and data augmentation.

- Retrospective data collection involves collecting data from past medical records or other sources. This data can be used to train and test the prediction model, and can be useful for identifying risk factors and other predictors of liver disease.
- Prospective data collection involves collecting new data from patients as they are diagnosed with liver disease. This data can be used to validate the model and ensure that it remains accurate over time.
- Data augmentation involves using additional data sources to improve the accuracy of the prediction model. For example, external data sources such as public health databases or social determinants of health data can be incorporated to help identify additional risk factors and predictors of liver disease.

### **Challenges in Data Collection:**

- There are several challenges associated with data collection for liver disease prediction. One challenge is the need for large amounts of data to train and validate the model. This can be particularly challenging for rare liver diseases, where there may be limited data available.

- Another challenge is the need for accurate and comprehensive data. Inaccurate or incomplete data can lead to errors and inaccuracies in the prediction model. Additionally, there may be variability in how data is collected across different healthcare providers, which can impact the accuracy of the model.

## **DATA PRE-PROCESSING**

The first step in training an SVM model is to pre-process the data. This includes cleaning the data, handling missing values, and scaling the features. Cleaning the data involves removing any outliers or errors that could affect the accuracy of the model. Handling missing values can be done by either removing the corresponding data points or imputing the missing values. Scaling the features is important to ensure that all features have the same scale and are equally important in the model.

## **FEATURE EXTRACTION**

Feature extraction is the process of selecting and transforming the raw data into a set of features that can be used to train a machine learning model for liver disease prediction. The goal of feature extraction is to identify and select the most relevant features that have the most predictive power for the target variable, which in this case is the presence or absence of liver disease.

The raw data used for liver disease prediction can come from a variety of sources, including electronic health records, medical imaging studies, and laboratory tests. The raw data may contain a large number of variables, which can make it difficult to train an accurate machine learning model. Feature extraction is therefore an important step to help reduce the dimensionality of the data and focus on the most relevant variables.

There are several methods used for feature extraction in liver disease prediction, including:

- **Principal component analysis (PCA):** PCA is a technique used to transform a large number of variables into a smaller number of uncorrelated variables, known as principal components. The principal components are ordered based on their variance, and the first few components with the highest variance are selected as the features.
- **Recursive feature elimination (RFE):** RFE is a technique that uses a machine learning model to select the most important features. The model is trained on the full set of features, and the least important features are removed one at a time until the optimal set of features is identified.
- **SelectKBest:** SelectKBest is a method that selects the K best features based on a statistical test. The most common statistical tests used include chi-squared, ANOVA F-test, and mutual information.
- **Domain knowledge:** In some cases, domain knowledge can be used to identify the most relevant features for liver disease prediction. For example, if a particular laboratory test is known to be a strong predictor of liver disease, it can be selected as a feature.

Once the features have been selected, they are transformed into a format that can be used to train a machine learning model, such as a matrix or a set of vectors. The transformed features are then used to train and test the machine learning model.

## MODEL TRAINING

Support Vector Machine is a powerful machine learning algorithm that is widely used for classification tasks, including liver disease prediction. SVM algorithm works by finding the optimal hyperplane that separates the different classes of data points. In liver disease prediction, the SVM algorithm is trained on a

dataset of features that are extracted from various sources such as electronic health records, medical imaging studies, and laboratory tests.

The SVM model is trained on the training dataset using the selected features. The SVM algorithm works by finding the optimal hyperplane that separates the different classes of data points. This hyperplane is selected based on the maximum margin, which is the distance between the hyperplane and the closest data points.

## **MODEL TESTING**

Once the SVM model is trained, it is evaluated on the testing dataset to measure its performance. The most common evaluation metrics used for SVM models include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC).

The final step in training an SVM model is to tune the hyperparameters to optimize the performance of the model. Hyperparameters such as the kernel type, the kernel coefficient, and the regularization parameter can be tuned using techniques such as grid search or random search.

## **2.2 SYSTEM DESIGN AND IMPLEMENTATION**

### **2.2.1 ALGORITHMS**

#### **2.2.1.1 SVM ALGORITHM**

SVM algorithm tries to give out hyper planes and split the data into different categories. The scikit-learn package in python is employed for implementing SVM. The pre-processed information is split into check information and coaching set that is of twenty fifth and seventy fifth of the entire dataset severally. A SVM technique builds hyper planes in an exceedingly dimensional area. a decent separation is achieved by the hyper plane that has the most important distance to the closest



coaching information of any category (so-called purposeful margin), since generally the larger the margin the lower the generalization error of the classifier.

#### **2.2.1.2 KNN ALGORITHM**

KNN is algorithm which is also called as the K-nearest neighboring algorithm. It is also called as the lazy learning algorithm. Its motivation is to utilize a database in which the information focuses are isolated into a few classes to anticipate the order of another example point.

#### **2.2.1.3 RANDOM FOREST**

Random Forest is also called as Random Decision Trees. Random Forest algorithm is a machine learning technique where these are a learning tasks, classification and regression tasks.

#### **2.2.1.4 DECISION TREE**

A decision tree is a flowchart-like structure in which each interior hub speaks to a "test" on a characteristic (for example, regardless of whether a coin flip comes up heads or tails), each branch speaks of the result of the test, and each leaf node speaks to a class mark (choice taken in the wake of processing all qualities). The ways from root to leaf speak to characterization rules.

#### **2.2.1.5 GRADIENT BOOSTER**

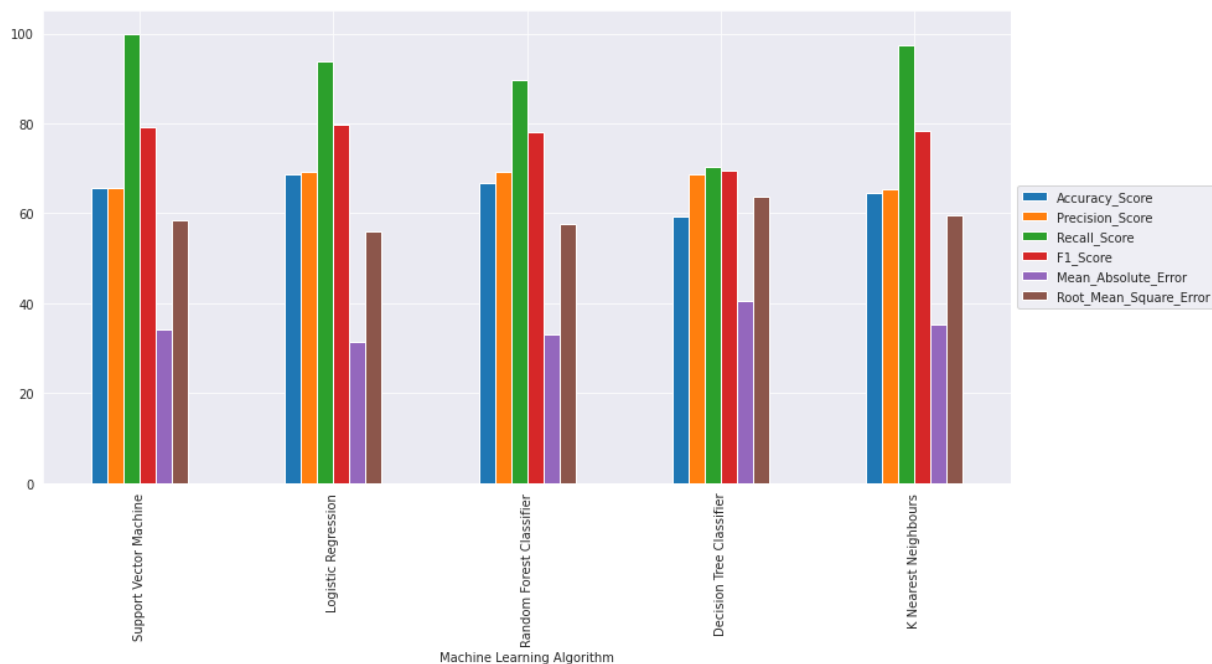
Gradient boosting is an AI system for relapse and order issues, which delivers an expectation show as a group of feeble forecast models, commonly decision trees. It manufactures the model in a phase savvy style like other boosting strategies do, and it sums them up by permitting enhancement of a self-assertive differentiable lost function.

## 2.3 ALGORITHM IMPLEMENTATION

Classification is the major data mining technique which is primarily used in healthcare sectors for medical diagnosis and predicting diseases. This project work used classification algorithms namely Support Vector Machine (SVM) for liver disease prediction.

Comparisons of different algorithms are done and it is based on the performance factors classification accuracy and execution time. From the experimental results, we concluded that SVM classifier is considered as the best.

### ANALYZING THE DIFFERENT ML ALGORITHMS:



**Figure 4. Machine Learning Algorithms**

## **2.4 TYPES OF TESTING**

### **2.4.1 UNIT TESTING**

In the development of a liver disease prediction model, unit testing can be used to ensure that each individual component of the model is working correctly and producing the expected outputs. This can include testing the data preprocessing steps, feature selection and engineering, model training, and model evaluation. For example, during data preprocessing, unit testing can be used to ensure that the data cleaning and transformation steps are producing the expected results. This can involve testing the removal of missing values, handling of outliers, and normalization of data.

In the feature selection and engineering step, unit testing can be used to ensure that the selected features are relevant to the problem and the engineered features are created correctly. In the model training step, unit testing can be used to ensure that the model is being trained correctly, and the hyperparameters are optimized correctly. This can involve testing the loss function and optimizer, training and validation accuracy, and regularization.

In the model evaluation step, unit testing can be used to ensure that the model is performing as expected and meeting its requirements. This can involve testing the confusion matrix, ROC curve, AUC, and other performance metrics. By conducting unit testing at each stage of the liver disease prediction model development, developers can identify and fix errors and ensure that the model is working as intended. This can help to improve the quality of the model and reduce the risk of errors and inaccuracies in the final prediction.

### **2.4.2 ALPHA-BETA TESTING**

Alpha testing is a type of software testing where the software is tested in a

simulated environment before it is released to the public. In the context of liver disease prediction, alpha testing would involve testing the model internally within the organization or research team to identify and fix any issues or bugs. It can involve multiple stages, including unit testing, integration testing, and system testing. Unit testing and integration testing are used to test individual components of the model and ensure that they are working correctly. System testing is used to test the entire system as a whole and ensure that all the components are working together as intended.

Once the liver disease prediction model has been successfully tested internally, it can then move on to beta testing. Beta testing is a type of software testing where the software is released to a select group of users outside of the development team to test the software in a real-world environment. In the context of liver disease prediction, beta testing would involve releasing the model to a group of users or patients who have agreed to participate in the testing. During beta testing, users can provide feedback on the performance of the model, including any issues or errors that they encounter. This feedback can be used to make further improvements to the model before it is released to the public.

By conducting alpha and beta testing, developers can ensure that the liver disease prediction model is working correctly and meeting the requirements of the users. This can help to improve the quality of the model and reduce the risk of errors and inaccuracies in the final prediction.

### **2.4.3 INTEGRATION TESTING**

Integration testing is a type of software testing where individual components of a system are tested together to ensure that they work correctly as a whole. In the context of liver disease prediction, integration testing would involve testing the

individual components of the prediction model together to ensure that they are working correctly and producing accurate results. Integration testing can be conducted after each component of the model has been tested separately using unit testing. The goal of integration testing is to ensure that the components work together as expected and produce accurate predictions.

In a liver disease prediction model, integration testing could involve testing the data preprocessing steps, feature selection and engineering, model training, and model evaluation together. This can involve testing how the feature selection and engineering steps impact the model's performance, or how different hyperparameter values impact the model's accuracy. Integration testing can also help to identify any issues or bugs that may arise when individual components are combined. By conducting integration testing, developers can identify and fix these issues before the model is deployed in a real-world setting.

By ensuring that all the components of the liver disease prediction model work together correctly, integration testing can help to improve the quality of the model and reduce the risk of errors and inaccuracies in the final prediction.

#### **2.4.4 ACCEPTANCE TESTING**

Acceptance testing is a type of software testing where the software is tested to determine whether it meets the requirements and specifications of the end-user. In the context of liver disease prediction, acceptance testing would involve testing the prediction model with real patient data to determine whether it meets the needs of healthcare professionals and patients. The goal of acceptance testing is to ensure that the prediction model is accurate and reliable, and that it provides useful information that can be used to make decisions about the diagnosis and treatment of liver disease. Acceptance testing can involve a variety of tests, including functional

testing, usability testing, and performance testing.

Functional testing is used to ensure that the model works correctly and provides accurate predictions for a variety of inputs. Usability testing is used to evaluate the ease of use of the model and ensure that it is accessible and understandable to healthcare professionals and patients. Performance testing is used to evaluate the speed and accuracy of the model, as well as its ability to handle large volumes of data.

During acceptance testing, the model may be tested using real patient data, which can be obtained from electronic health records or other sources. This can help to identify any issues or inaccuracies in the model, which can then be addressed before the model is released for clinical use.

By conducting acceptance testing, developers can ensure that the liver disease prediction model meets the needs of healthcare professionals and patients, and that it provides accurate and reliable information that can be used to improve patient outcomes.

## CHAPTER 3

### CONCLUSION

#### 3.1 CONCLUSION

Through this project we have increased the efficiency of the prediction. We have increased the accuracy of the prediction algorithms where we have used different algorithms to predict the accuracy of the disease at different accuracy levels.

Classification is the major data mining technique which is primarily used in healthcare sectors for medical diagnosis and predicting diseases. This project work used classification algorithms namely Support Vector Machine (SVM) for liver disease prediction.

Comparisons of different algorithms are done and it is based on the performance factors classification accuracy and execution time. From the experimental results, we concluded that SVM classifier is considered as the best.

We have used a specific dataset Indian liver patient dataset where we have 10 attributes and more than 500 patients data so it would be very useful and give best accuracy of the prediction.

#### 3.2 FUTURE ENHANCEMENT

- **Integration of genetic data:** Currently, most machine learning models for liver disease prediction rely on clinical and demographic variables. However, genetic data could also provide valuable information for predicting the risk of liver disease.
- **Real-time monitoring:** Machine learning models could be integrated with wearable devices or sensors to provide real-time monitoring of liver function and detect changes in liver health over time.
- **Improved accuracy:** Although machine learning models have shown promising results for liver disease prediction, there is still room for

improvement in terms of accuracy. Future research could focus on developing more accurate models that can better differentiate between different types of liver disease.

- **Personalized risk assessment:** Machine learning models could be used to provide personalized risk assessments for liver disease based on an individual's unique medical history and lifestyle factors.
- **Integration with electronic health records:** Machine learning models could be integrated with electronic health records to improve the accuracy of predictions and provide clinicians with real-time information on patients' liver health.

## APPENDICES

### (A) SOURCECODE

#### MODEL BUILDING

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

#### READING THE DATASET

```
data=pd.read_csv("indian_liver_patient.csv")
```

#### EXPLORATORY DATA ANALYSIS

```
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Age                   583 non-null   int64
1   Gender                583 non-null   object
2   Total_Bilirubin       583 non-null   float64
3   Direct_Bilirubin      583 non-null   float64
4   Alkaline_Phosphotase  583 non-null   int64
```



```

5  Alamine_Aminotransferase  583 non-null  int64
6  Aspartate_Aminotransferase 583 non-null  int64
7  Total_Protiens           583 non-null  float64
8  Albumin                  583 non-null  float64
9  Albumin_and_Globulin_Ratio 579 non-null  float64
10 Dataset                  583 non-null  int64
dtypes: float64(5), int64(5), object(1)
memory usage: 50.2+ KB

```

## DATA VISUALIZATION

### Univariate Analysis

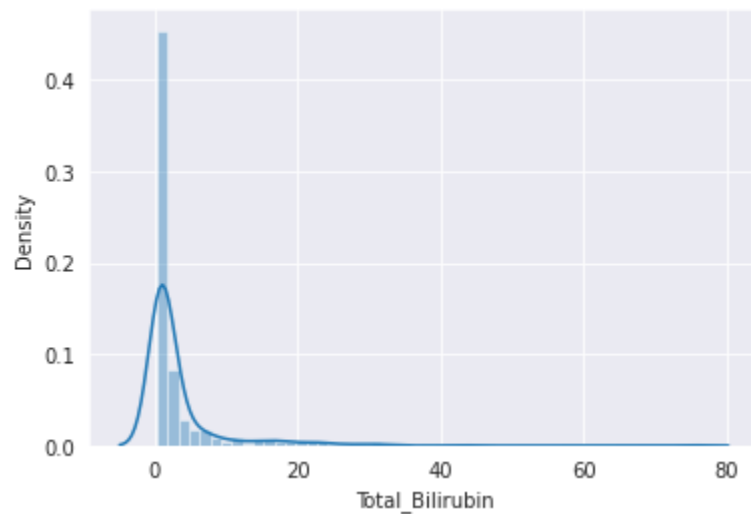
```

import warnings
warnings.filterwarnings("ignore")

```

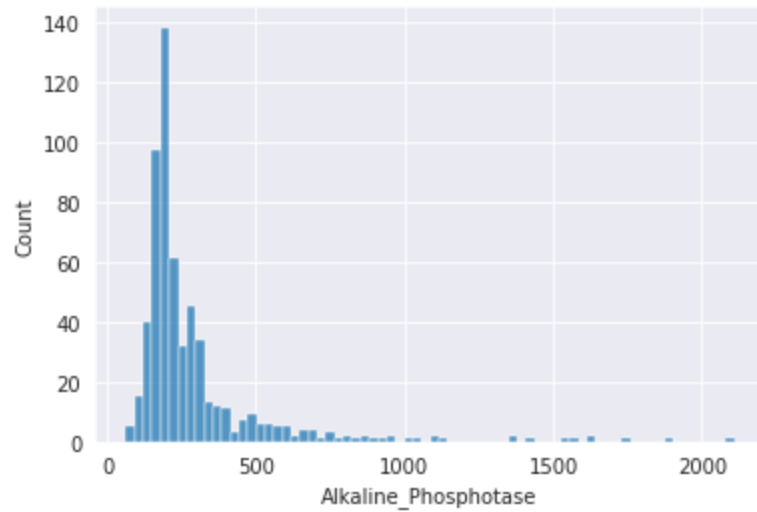
```
sns.distplot(data["Total_Bilirubin"])
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fc346919590>



```
sns.histplot(data["Alkaline_Phosphotase"])
```

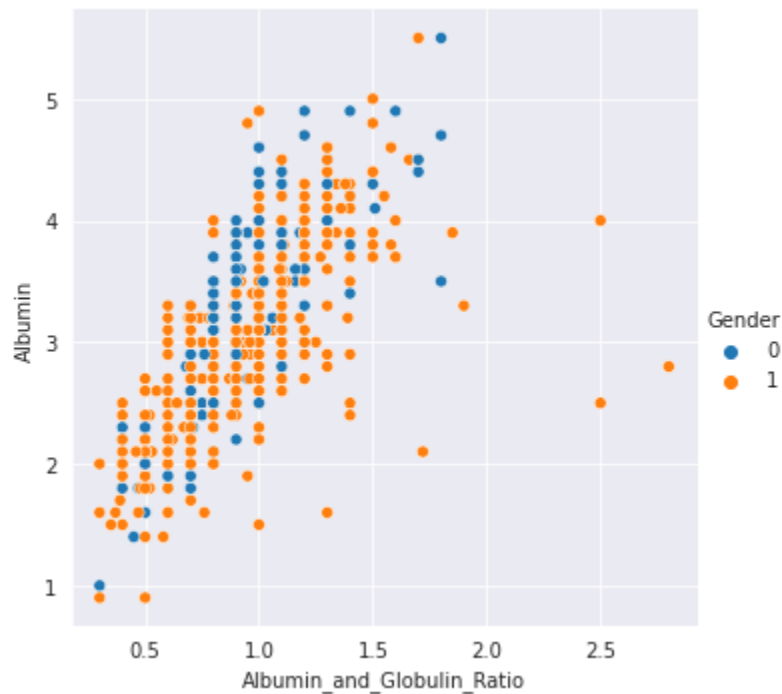
<matplotlib.axes.\_subplots.AxesSubplot at 0x7fc3462a5610>



## Bivariate Analysis

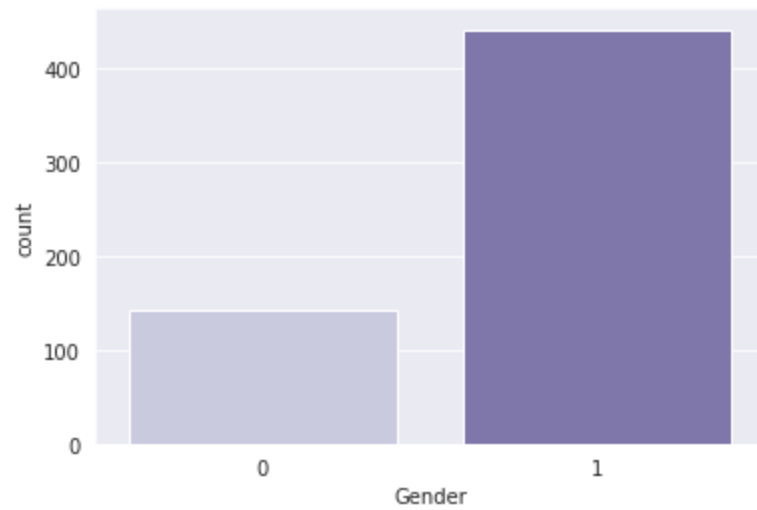
```
sns.relplot(x="Albumin_and_Globulin_Ratio",y="Albumin",data=data,hue="Gender")
```

<seaborn.axisgrid.FacetGrid at 0x7fc3460b2150>



## Multi - Variate Analysis

```
sns.countplot(data.Gender,palette='Purples')
plt.show()
```



```
sns.catplot(x="Gender",y="Direct_Bilirubin",data=data)
```

Out[257]:

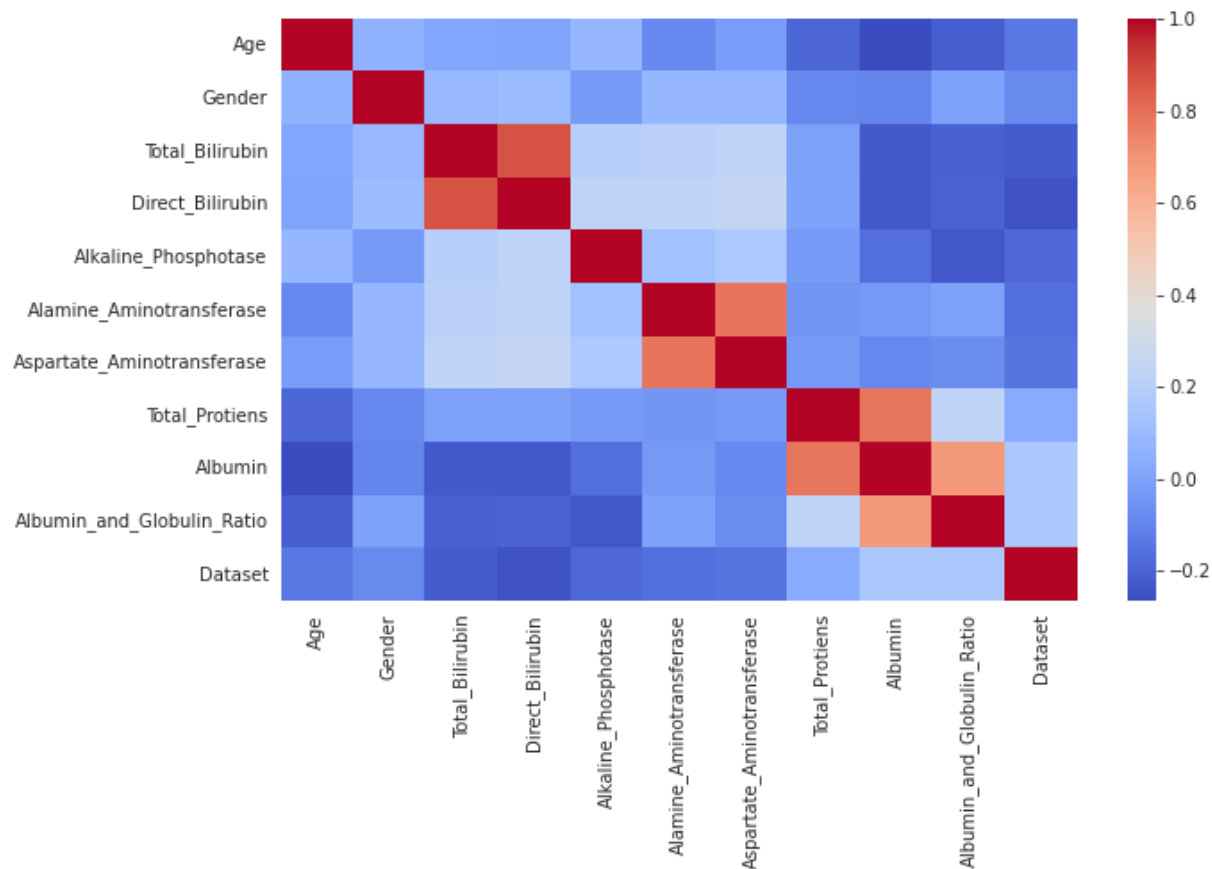
```
<seaborn.axisgrid.FacetGrid at 0x7fc3461baed0>
```



```
fig=plt.figure(figsize=(10,6))
```

```
sns.heatmap(data.corr(),cmap='coolwarm')
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fc3459e7a10>



## SPLITTING THE DATASET INTO INDEPENDENT AND DEPENDENT VARIABLE

In [267]:

```
X=liver_scaled
y=data['Dataset']
Splitting the independent and dependent features into train and test set
```

In [268]:

```
from sklearn.model_selection import train_test_split
from sklearn.metrics import precision_score,accuracy_score,recall_score,f1_score,mean_absolute_error,mean_squared_e
rror
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3)
y_test.value_counts()
```

Out[268]:

```
1    115
2     60
Name: Dataset, dtype: int64
```

## MODEL BUILDING

### SUPPORT VECTOR MACHINE

```
from sklearn.svm import SVC
model = SVC(kernel = 'rbf', random_state = 0)
model.fit(X_train, y_train)
```

Out[269]:

```
SVC(random_state=0)
```

In [270]:

```
svm_predictions = model.predict(X_test)
```

In [271]:

```
from sklearn.metrics import classification_report, confusion_matrix
```

In [272]:

```
svm_acc=accuracy_score(y_test,svm_predictions)
svm_prec=precision_score(y_test,svm_predictions)
svm_rec=recall_score(y_test,svm_predictions)
svm_f1=f1_score(y_test,svm_predictions)
svm_mas=mean_absolute_error(y_test,svm_predictions)
svm_rms=np.sqrt(mean_squared_error(y_test,svm_predictions))
```

In [273]:

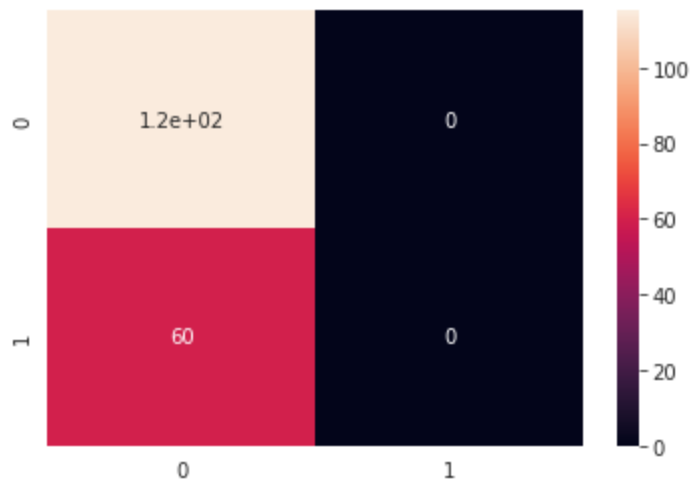
```
scm=confusion_matrix(y_test,svm_predictions)
```

In [274]:

```
sns.heatmap( scm, annot=True)
```

Out[274]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fc3458c7750>



## SAVE THE MODEL

In [317]:

```
import pickle
pickle.dump(model, open('SVM.pkl','wb'))
```

## HTML & CSS

### Index.html

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

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

</head>
<body>
<div class="b">
  <h2 class="body_heading"><span class="heading">Liver Disease Prediction
Model</span></h2>
</div>

<div class="division">
  <form action="{{ url_for('predict') }}" method="POST">
    <br>
    <div class="d1">
      <h3>Age</h3>
      <input id="age" name="Age" placeholder="in Year" required="required">
      <br>
      <h3>Gender</h3>
      <input id="gender" name="Gender" placeholder="Male/Female" required="required">
      <br>
      <h3>Total Bilirubin</h3>
      <input id="bil" name="Total_Bilirubin" placeholder="Total Bilirubin"
required="required">
      <br>
      <h3>Direct Bilirubin</h3>
      <input id="dir" name="Direct_Bilirubin" placeholder="Direct Bilirubin"
required="required">
      <br>
      <h3>Alkaline Phosphatase</h3>
      <input id="alk" name="Alkaline_Phosphatase" placeholder="Alkaline Phosphatase"
required="required">
    </div>
    <br>
    <div class="d2">
      <h3>Alamine Aminotransferase</h3>
      <input id="ala" name="Alamine_Aminotransferase" placeholder="Alamine
Aminotransferase" required="required">
    </div>
  </form>
</div>
```

```

        <br>
        <h3>Aspartate Aminotransferase</h3>
        <input id="asp" name="Aspartate_Aminotransferase" placeholder="Aspartate
Aminotransferase" required="required">
        <br>
        <h3>Total Proteins</h3>
        <input id="tot" name="Total_Proteins" placeholder="Total Proteins"
required="required">
        <br>
        <h3>Albumin</h3>
        <input id="alb" name="Albumin" placeholder="Albumin" required="required">
        <br>
        <h3>Albumin and Globulin Ratio</h3>
        <input id="glo" name="Albumin_and_Globulin_Ratio" placeholder="Albumin and
Globulin Ratio" required="required">
    </div>
    <br>
    <br>
    <br>
    <button id="sub" type="submit">Submit</button>
</form>
</div>
</body>
</html>

```

## **Result.html**

```

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

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

<body>

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

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

```

```

        <div class="result">
            {% if prediction==2 %}
                <h1><span class='alert'>Oh no! <br><br>You have Liver disease <br><br>Please do
consult a doctor for more information</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>

    <p class='footer-description'>Health is wealth</p>

</div>
</body>
</html>

```

## FLASK APP

### FlaskApp.py

```

from flask import Flask, render_template, request
import numpy as np
import pickle
import sklearn

```

```

app = Flask(__name__)
model = pickle.load(open('SVM.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 = request.form['Gender']

```



```

Total_Bilirubin = float(request.form['Total_Bilirubin'])
Direct_Bilirubin = float(request.form['Direct_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,Direct_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)

```

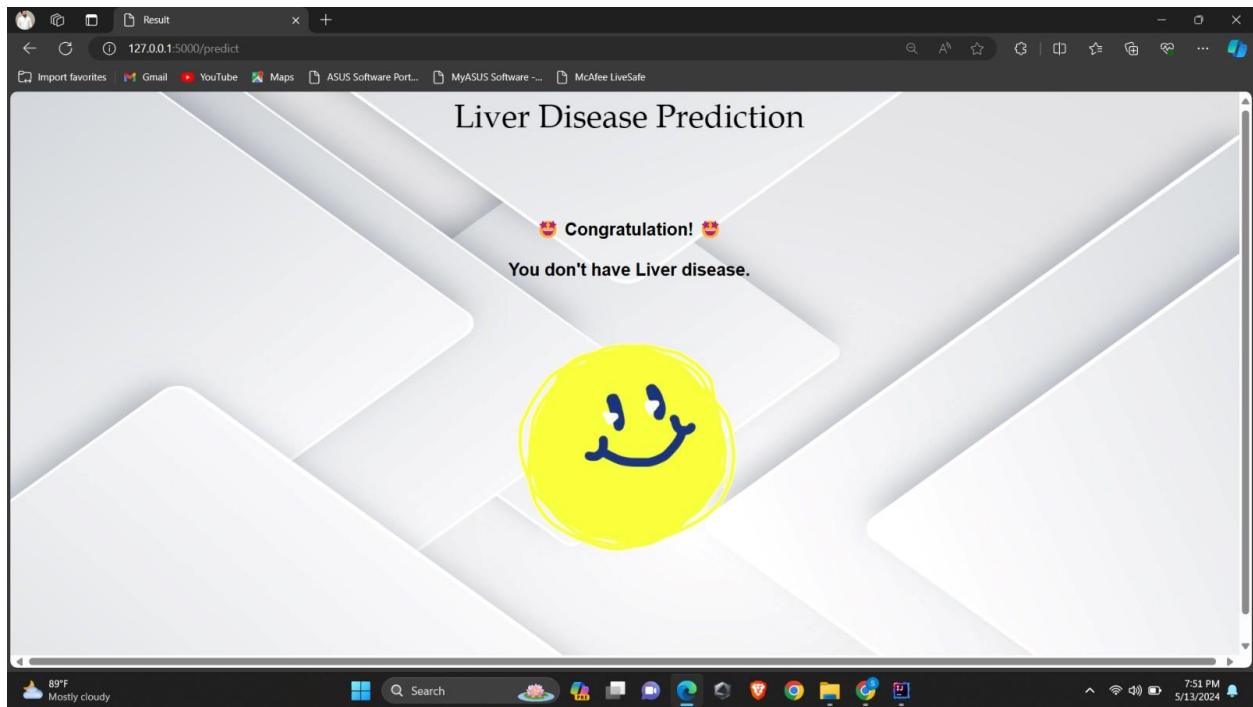
## (B) SNAPSHOTS

**Liver Disease Prediction Model**

Age	12	Alamine Aminotransferase	23
Gender	1	Aspartate Aminotransferase	33
Total Bilirubin	0.7	Total Protiens	6.6
Direct Bilirubin	0.9	Albumin	3.9
Alkaline Phosphotase	185	Albumin and Globulin Ratio	9.00

Submit

**Figure 5. Input page**



**Figure 6. Output Page**

## REFERENCES

1. Alzahrani, A., et al. "An efficient approach for liver disease prediction using machine learning techniques." 2020 11th International Conference on Information and Communication Systems (ICICS). IEEE, 2020.
2. Bora, Jyotishmita, et al. "Liver disease prediction using machine learning algorithms: a systematic review." Journal of Medical Systems 43.8 (2019): 238.
3. Choudhary, Preeti, et al. "A survey on liver disease prediction using machine learning techniques." International Journal of Advanced Research in Computer Science and Software Engineering 8.6 (2018): 177-183.
4. Gajananan, K. S., et al. "Liver disease prediction using machine learning algorithms." 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.
5. Naeem, Muhammad, et al. "Liver disease prediction using machine learning algorithms: a review." Journal of Medical Systems 45.7 (2021): 1-16.
6. Pandey, Anand, and Nidhi Sharma. "Liver disease prediction using machine learning algorithms." International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) 8.6 (2019): 9-12.
7. Sharma, Mukesh, et al. "Liver disease prediction using machine learning algorithms." 2021 International Conference on Communication, Computing and Electronics Systems (ICCCES). IEEE, 2021.
8. Varghese, Asha, and S. S. Lakshmi. "Liver disease prediction using machine learning techniques." 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). IEEE, 2017.
9. Zhang, Huan, et al. "Prediction of liver disease using machine learning algorithms based on clinical variables." Journal of healthcare engineering 2021 (2021).
10. Zhang, Guodong, et al. "Liver disease prediction using ensemble machine learning." 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2019.