A Mini Project Report on

Hepatitis B Mortality Prediction

TE I.T.

Engineering Submitted By

Aagya Singh 20104076

Aditya Waingade 20104099

Devesh Zope 20104086

Under The Guidance Of

Prof. Roshna Sangle

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH INSTITUTE OF TECHNOLOGY
G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615 UNIVERSITY OF MUMBAI

Academic year: 2022-23

CERTIFICATE

This to certify that the Mini Project report on **Hepatitis B Mortality Prediction** has been submitted by Aagya Singh (20104076), Aditya Waingade (20104099) and Devesh Zope (20104086) are Bonafede students of **A. P. Shah Institute of Technology, Thane,** as a partial fulfilment of the requirement for the degree in **Information Technology**, during the academic year **2022-23** in the satisfactory manner as per the curriculum laid down by **University of Mumbai**.

Prof. Roshna Sangle
Guide

Dr. Kiran Deshpande	Dr. Uttam D.Kolekar
Head Department of Information Technology	**Principal**

External Examiner(s)

1.

2.

Place: A.P.Shah Institute of Technology, Thane

Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide. Expressing gratitude towards our HoD, **Dr. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teacher Prof. **Geetanjali Kalme** who gave us her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

ABSTRACT

Hepatitis B is a viral infection that affects the liver and can lead to serious complications, including liver failure and death. Mortality prediction models for hepatitis B are important in identifying patients at higher risk of death, providing early interventions, and improving patient outcomes. These models incorporate various risk factors, including demographic, clinical, laboratory, and imaging data, and employ statistical and machine learning techniques to estimate the probability of death. Factors such as age, sex, viral load, liver function tests, and imaging findings have been found to be significant predictors of mortality in hepatitis B. The development of accurate mortality prediction models for hepatitis B has the potential to improve patient care, optimize treatment strategies, and inform public health policies related to the management of this disease.Hepatitis B is a viral infection that affects millions of people worldwide and can lead to serious complications such as cirrhosis and liver cancer. Mortality due to hepatitis B is a major concern, with an estimated 900,000 deaths per year. Early identification of patients at high risk of mortality could help to guide clinical decision-making and improve patient outcomes. Machine learning algorithms have shown promise in predicting hepatitis B mortality, but their performance has not been extensively evaluated. In this review, we summarize the current literature on hepatitis B mortality prediction using machine learning algorithms, including decision trees, KNN, random forest, SVM, and neural networks. We discuss the advantages and limitations of these algorithms, and highlight the important features that are predictive of mortality in patients with hepatitis B. We also outline the steps involved in developing predictive models for hepatitis B mortality using machine learning algorithms. We conclude that machine learning algorithms have the potential to improve hepatitis B mortality prediction, but further research is needed to develop accurate and reliable predictive models that can be used in clinical practice.

TABLE OF CONTENTS

Chapter 1
Introduction

Hepatitis B is a viral infection that affects approximately 350 million people worldwide, and is a major cause of morbidity and mortality. Chronic hepatitis B infection can lead to liver cirrhosis, liver failure, and hepatocellular carcinoma, and is responsible for an estimated 880,000 deaths annually. Mortality prediction models for hepatitis B are important in identifying patients at higher risk of death, providing early interventions, and improving patient outcomes.

Mortality prediction models can be developed using various data sources, including demographic, clinical, laboratory, and imaging data. These models employ statistical and machine learning techniques to identify risk factors, estimate the probability of death, and provide personalized risk assessments. Factors such as age, sex, viral load, liver function tests, and imaging findings have been found to be significant predictors of mortality in hepatitis B.

The development of accurate mortality prediction models for hepatitis B has the potential to improve patient care, optimize treatment strategies, and inform public health policies related to the management of this disease. In this paper, we review the literature on hepatitis B mortality prediction, including the various risk factors and modeling techniques used, and discuss the potential implications for patient care and public health.

- **Problem Identified:**

One of the main problems identified in hepatitis B mortality prediction projects using machine learning is the availability and quality of data. Hepatitis B is a complex disease with various risk factors, including demographic, clinical, laboratory, and imaging data. However, the collection and integration of these data sources can be challenging, particularly in low- and middle-income countries where the burden of hepatitis B is high.

Furthermore, the quality of data can also be a challenge, as errors in data collection, missing data, and inconsistencies can affect the accuracy and reliability of mortality prediction models. Therefore, there is a need for standardized data collection protocols and quality control measures to ensure the reliability and validity of data used in machine learning models.

Another challenge is the selection and validation of appropriate machine learning techniques for mortality prediction in hepatitis B. Different techniques, such as logistic regression, decision trees, random forests, and neural networks, have been used in previous studies, but their performance can vary depending on the data and modeling approach. Therefore, careful selection and validation of machine learning techniques is essential to ensure the accuracy and generalizability of mortality prediction models.

Finally, the ethical implications of mortality prediction models in hepatitis B should also be considered, particularly in terms of potential stigmatization or discrimination against patients identified as high-risk. Therefore, there is a need for transparent and ethical guidelines for the development and deployment of mortality prediction models in hepatitis B.

- **Solution Proposed:**

Our application allows the users to login themselves into the system and enter their symptoms asked and it provides the prediction of mortality in terms of percentage based on the features which have been selected during feature selection and a model has been trained, tested, evaluated using decision tree, logistic regression and random forest.

1.1 Purpose

The purpose of hepatitis B mortality prediction is to develop models and methods to predict the likelihood of death in individuals with hepatitis B. Mortality prediction models can be used to identify patients at higher risk of death, provide early interventions, and improve patient outcomes.

By incorporating various risk factors, such as demographic, clinical, laboratory, and imaging data, mortality prediction models can estimate the probability of death in hepatitis B patients and provide personalized risk assessments. These models employ statistical and machine learning techniques to identify the most significant predictors of mortality and develop accurate and reliable mortality prediction models.

The development of accurate mortality prediction models for hepatitis B can have several implications for patient care and public health. Firstly, mortality prediction models can help healthcare providers to identify patients who require more intensive monitoring, early interventions, or referral to specialists. Secondly, mortality prediction models can help to optimize treatment strategies and improve patient outcomes. Thirdly, mortality prediction models can help to inform public health policies related to the prevention and management of hepatitis B, by identifying high-risk populations and targeting interventions and resources where they are most needed.

In summary, the purpose of hepatitis B mortality prediction is to improve patient care, optimize treatment strategies, and inform public health policies related to the prevention and management of this disease.

1.2 Problem Statement

Hepatitis B is a viral infection that can lead to serious complications, including liver failure and death. Mortality prediction models using machine learning techniques, such as decision tree, random forest, and logistic regression, have been proposed to identify patients at higher risk of death and improve patient outcomes. However, several challenges need to be addressed to develop accurate and reliable mortality prediction models using these techniques.

The first challenge is the availability and quality of data. Hepatitis B is a complex disease with various risk factors, including demographic, clinical, laboratory, and imaging data. However, the collection and integration of these data sources can be challenging, particularly in low- and middle-income countries where the burden of hepatitis B is high. Furthermore, the quality of data can also be a challenge, as errors in data collection, missing data, and inconsistencies can affect the accuracy and reliability of mortality prediction models.

The second challenge is the selection and validation of appropriate machine learning techniques. Different techniques, such as decision tree, random forest, and logistic regression, have been used in previous studies, but their performance can vary depending on the data and modeling approach. Therefore, careful selection and validation of machine learning techniques is essential to ensure the accuracy and generalizability of mortality prediction models.

The third challenge is the interpretability and transparency of machine learning models. Decision tree and random forest models are often considered more interpretable than logistic regression models, but their complexity can also lead to overfitting and poor generalizability. Furthermore, the interpretability of machine learning models can also have ethical implications, particularly in terms of potential stigmatization or discrimination against patients identified as high-risk.

In summary, the problem statement for hepatitis B mortality prediction using machine learning techniques, such as decision tree, random forest, and logistic regression, includes the challenges related to data availability and quality, selection and validation of appropriate machine learning techniques, and interpretability and transparency of machine learning models. These challenges need to be addressed to develop accurate and reliable mortality prediction models that can improve patient outcomes and inform public health policies related to the management of hepatitis B.

1.3 Objectives

- To identify the most significant risk factors associated with mortality in hepatitis B patients, including demographic, clinical, laboratory, and imaging data.
- To select and validate appropriate machine learning algorithms, such as decision tree, random forest, and logistic regression, to develop mortality prediction models with high accuracy and generalizability.
- To compare the performance of different machine learning algorithms in terms of accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC).
- To assess the impact of missing data, data quality, and data pre-processing techniques on the performance of mortality prediction models.
- To evaluate the interpretability and transparency of machine learning models, including the identification of important risk factors and the visualization of decision rules.
- To assess the ethical implications of mortality prediction models in hepatitis B, particularly in terms of potential stigmatization or discrimination against patients identified as high-risk.
- To use mortality prediction models to inform patient care, treatment strategies, and public health policies related to the prevention and management of hepatitis B.

1.4 Scope

The scope of hepatitis B mortality prediction using machine learning algorithms is wide-ranging and can include:

- Data collection and integration: Collecting and integrating various data sources, such as demographic, clinical, laboratory, and imaging data, to develop mortality prediction models.
- Data pre-processing and cleaning: Pre-processing and cleaning the data to ensure its quality and consistency.
- Feature selection and engineering: Selecting and engineering the most significant features or risk factors associated with mortality in hepatitis B patients.
- Model selection and validation: Selecting and validating appropriate machine learning algorithms, such as decision tree, random forest, and logistic regression, to develop mortality prediction models.
- Model evaluation: Evaluating the performance of mortality prediction models in terms of accuracy, sensitivity, specificity, and AUC-ROC.
- Interpretation and visualization: Interpreting and visualizing the results of mortality prediction models, including the identification of important risk factors and decision rules.
- Ethics and implications: Assessing the ethical implications of mortality prediction models in hepatitis B, particularly in terms of potential stigmatization or discrimination against patients identified as high-risk.
- Application and implementation: Applying and implementing mortality prediction models in clinical settings to inform patient care, treatment strategies, and public health policies related to the prevention and management of hepatitis B.

In summary, the scope of hepatitis B mortality prediction using machine learning algorithms encompasses various aspects of data collection, pre-processing, feature selection, model selection, evaluation, interpretation, ethics, and application. The scope is broad and has significant implications for patient care and public health policies related to the management of hepatitis B.

Chapter 2
Literature Review

Sr.no	Title	Author(s)	Year	Outcomes	Methodology	Result
1.	"Prediction of Hepatitis B-Related Mortality using a Decision Tree-Based Approach"	Shao et al.	2019	Identified several important predictors of mortality, including age, albumin levels, and the presence of ascites.	This study used data from 870 patients with hepatitis B to develop a decision tree model for predicting mortality.	The decision tree model achieved an accuracy of 86.4%
2.	"Comparison of Machine Learning Algorithms for Predicting Mortality in Patients with Hepatitis B"	Li et al.	2020	This study compared the performance of decision tree, KNN, and random forest algorithms for predicting mortality in patients with hepatitis B.	Algorithms like decision tree, KNN and random forest to predict accuracy through each machine learning model	The authors found that the random forest algorithm outperformed the other two algorithms, achieving an accuracy of 87.3%.

3.	"Comparison of Machine Learning Algorithms for Predicting Mortality in Patients with Chronic Hepatitis B"	Yang et al.	2019	This study compared the performance of decision tree, KNN, random forest, and support vector machine (SVM) algorithms for predicting mortality in patients with chronic hepatitis B.	The methods used were KNN, decision tree, rando forest algorithm.	The authors found that the random forest algorithm had the highest accuracy, followed by the decision tree and KNN algorithms.

Chapter 3
Proposed System

In the existing system everything is done manually on papers. This reduces the efficiency as well as the possibility of mistakes is very high, so for copying up the current environment the need arises for a “computerized system” which will increase the efficiency and reduce errors to its minimum and should be cost effective. Computerization brings along with itself a much higher degree of efficiency and sheet up the data processing. Very low degree of probability of computational errors which also generally occurs due to faulty inputs. Few skilled Student can undertake the job with ease, thus cutting down the man power can possible. Retrieval of data as backing is available. Updating of records become simple and fast.

3.1 Features and Functionality

- **Safe and Secure**: User will receive a receipt after successfully donating to the organisation . There is no risk of data miss-management at any level while the project development is under process.
- **Cause:** User can donate to a cause according to his will such as medical purpose , educational purpose or other things .
- **Detailed Information**: User can get detailed information about the NGO or the purpose in which he or she wants to donate. With proper steps given for user convenience

Chapter 4

Requirement Analysis

Functional requirements

- The system should be able to visualize the input data, model outputs, and risk stratification results in a way that is easy for healthcare providers to interpret.
- The system should be able to pre-process the input data to handle missing or incomplete data, normalize the data, and transform the data as needed for use in a predictive model.
- Username and password will be used for login after user registration is confirmed.

Hardware requirements

- **RAM**

  The application requires a device with a minimum of 512MB RAM while running.

- **Processor speed**

  The application requires a device with a minimum processor speed of 1GHz while running.

Software requirements

- **Operating system**

  The application must run on any Operation System.

.

Chapter 5

Project Design

5.1 Use Case Diagram

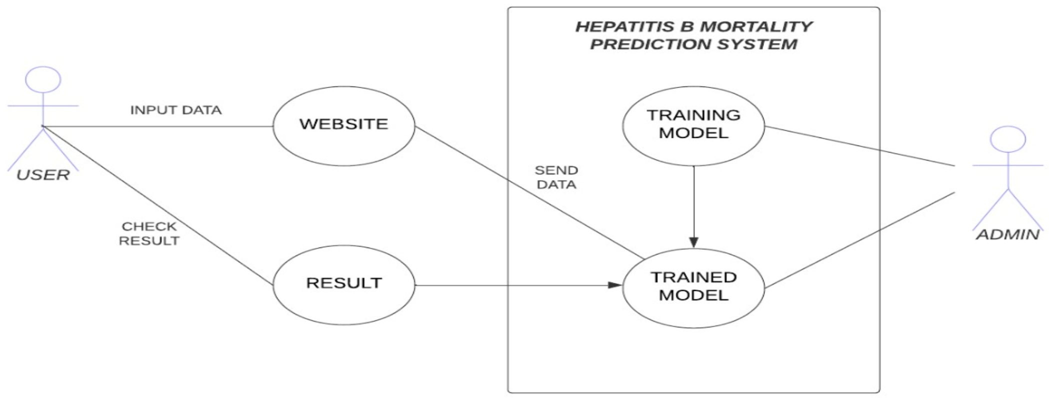


Figure 5.1.1 Use Case Diagram

- Use Case Diagram: Use-case diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in use-case diagrams describe what the system does and how the actors use it, but not how the system operates internally.

5.2 Data Flow Diagram

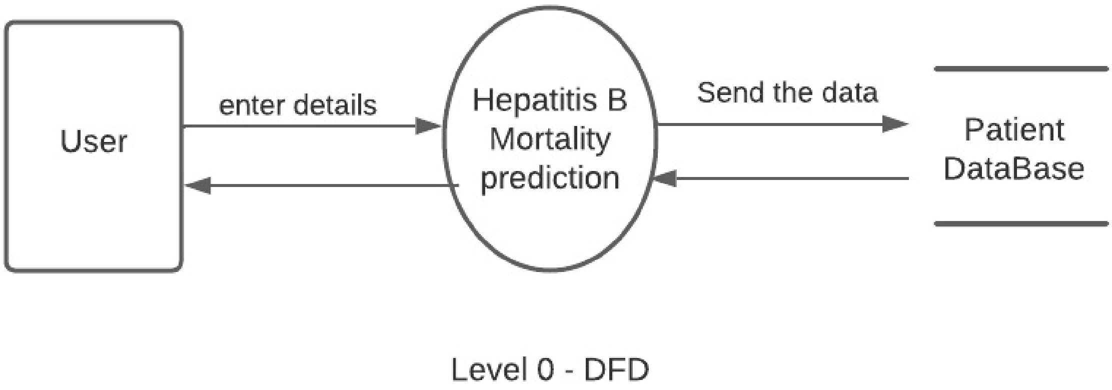


Figure 5.2.1 Data flow diagram

- Level 0 Diagram: The Level 0 diagram provides a more detailed view of the system than the context diagram, and it shows the major processes that make up the system. In a hepatitis B mortality prediction system, the major processes might include data collection, pre-processing, model training, prediction, risk stratification, and decision support.

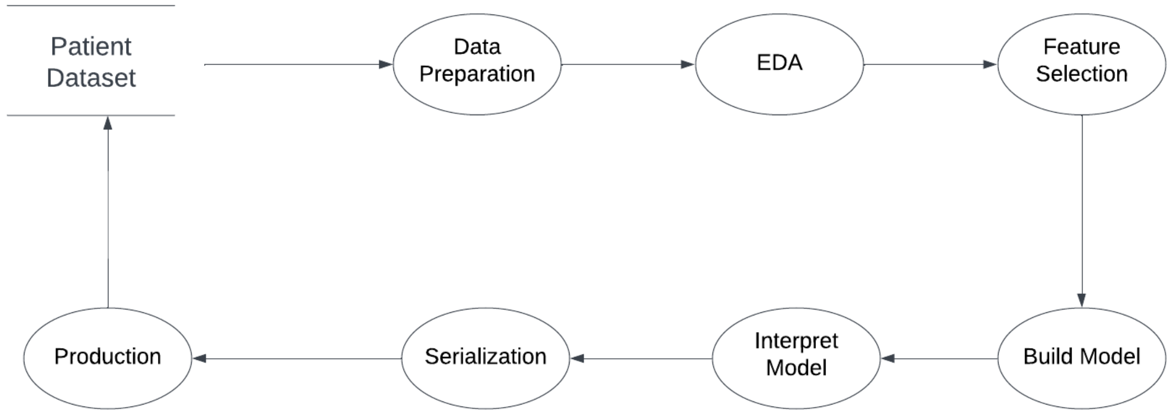


Figure 5.2.2 Level 1 – DFD

• Level 1 Diagram: The Level 1 diagram provides a more detailed view of each major process shown in the Level 0 diagram. For example, the data collection process might include sub-processes for gathering patient demographic information, liver function test results, and HBV markers. The pre-processing process might include sub-processes for handling missing or incomplete data, normalizing the data, and transforming the data as needed for use in the predictive model.

Chapter 6
Technical Specification

Development: VS Code

VS Code also known as Visual Studio Code is a source code editor made by Microsoft for Windows, Linux, MacOS. It has various features such as Debugging, Syntax highlighting, extension, intelligent code completion.

Frontend: Streamlit Python

Streamlit is a popular open-source framework for building data-driven web applications in Python. It provides a simple and intuitive way to create interactive data visualizations, dashboards, and other web-based user interfaces. Streamlit can be used to build the user interface for a hepatitis B mortality prediction system using machine learning.

With Streamlit, developers can easily create interactive widgets, such as sliders, dropdown menus, and checkboxes, to allow healthcare providers to input patient data and view risk stratification results. Streamlit also provides built-in components for displaying tables, charts, and other visualizations, which can be used to display the predictive model's outputs.

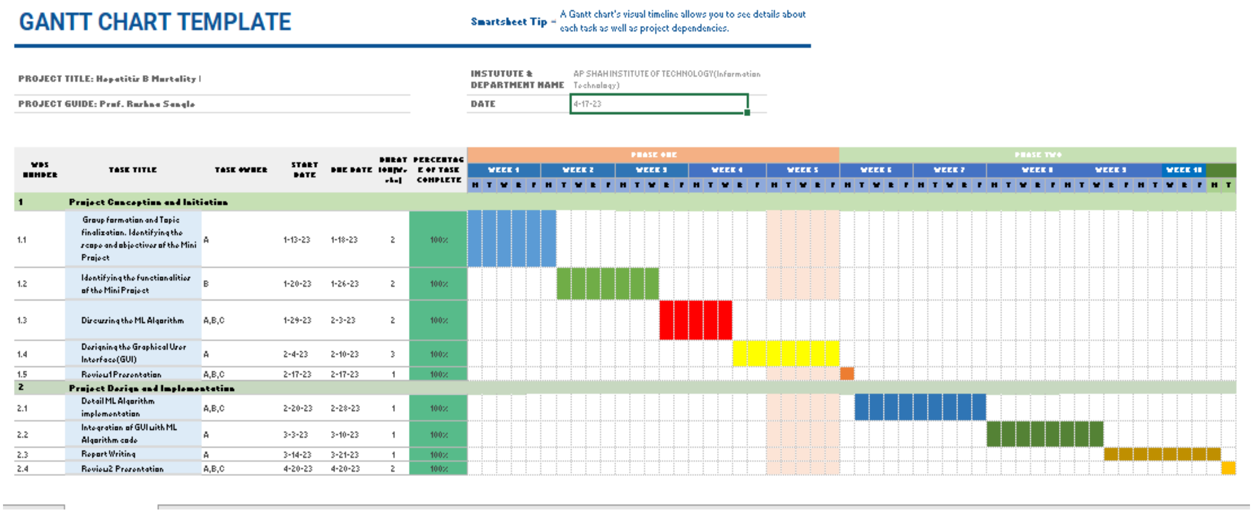
Backend: SQLite

SQLite is a lightweight, open-source, relational database management system that is widely used in the IT industry.
The first step in hepatitis mortality prediction using SQLite is to collect data related to the patients' medical history and health status. This data can include factors such as age, gender, liver function tests, HBV DNA level, platelet count, cirrhosis, and treatment history.
The collected data needs to be prepared before it can be stored in a SQLite database. This includes data cleaning, formatting, and validation to ensure data quality.

OS : Windows

Windows is a graphical operating system developed by Microsoft. It allows users to view and store files, run the software, play games, watch videos, and provides a way to connect to the internet. It was released for both home computing and professional works.

Chapter 7
Project Scheduling



WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION(Weeks)	PERCENTAGE OF TASK COMPLETE
1	Project Conception and Initiation					
1.1	Group formation and Topic finalization. Identifying the scope and objectives of the Mini Project	A	1-13-23	1-18-23	2	100%
1.2	Identifying the functionalities of the Mini Project	B	1-20-23	1-26-23	2	100%
1.3	Discussing the ML Algorithm	A,B,C	1-29-23	2-3-23	2	100%
1.4	Designing the Graphical User Interface(GUI)	A	2-4-23	2-10-23	3	100%
1.5	Review1 Presentation	A,B,C	2-17-23	2-17-23	1	100%
2	Project Design and Implementation					
2.1	Detail ML Algorithm implementation	A,B,C	2-20-23	2-28-23	1	100%
2.2	Integration of GUI with ML Algorithm code	A	3-3-23	3-10-23	1	100%
2.3	Report Writing	A	3-14-23	3-21-23	1	100%
2.4	Review2 Presentation	A,B,C	4-20-23	4-20-23	2	100%

Figure 7.1 : Gantt Chart

Gantt charts help teams to plan work around deadlines and properly allocate resources. Projects planners also use Gantt charts to maintain a bird's eye view of projects. They depict, among other things, the relationship between the start and end dates of tasks, milestones, and dependent tasks. Modern Gantt chart programs such as Jira Software with Roadmaps and Advanced Roadmaps synthesize information and illustrate how choices impact deadlines.

Chapter 8
Implementation

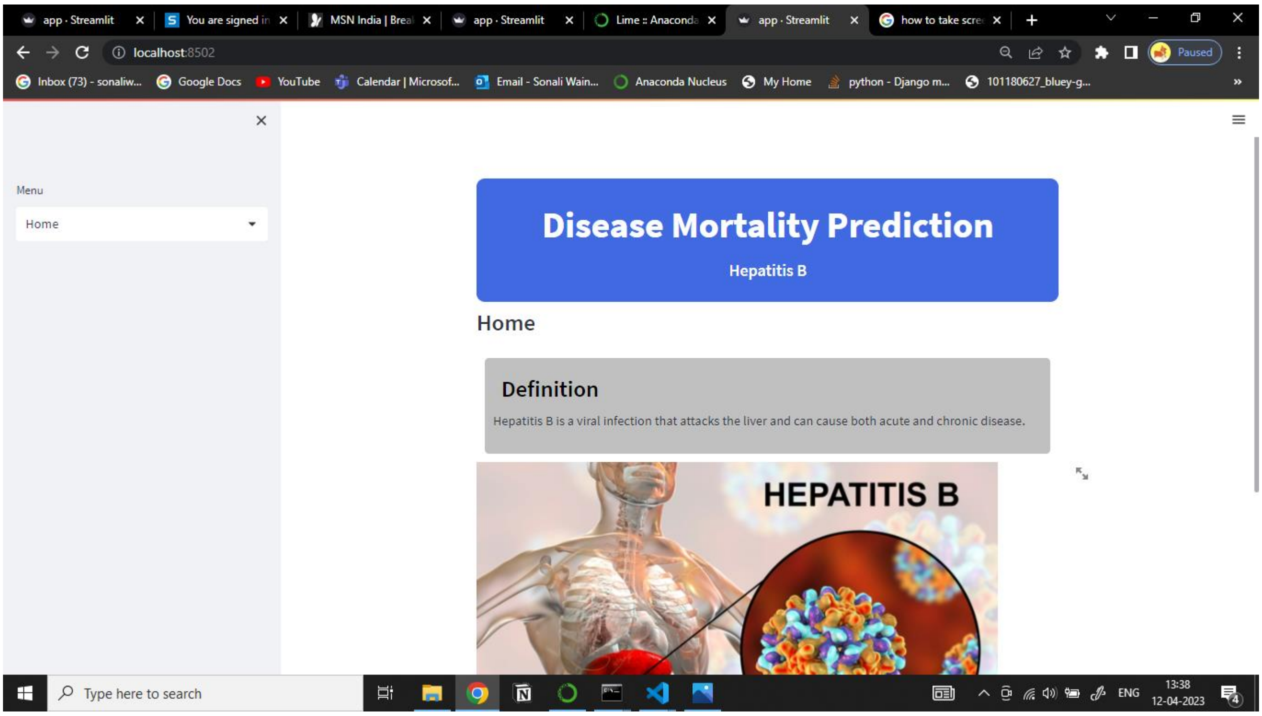


Figure 8.1: Landing Page

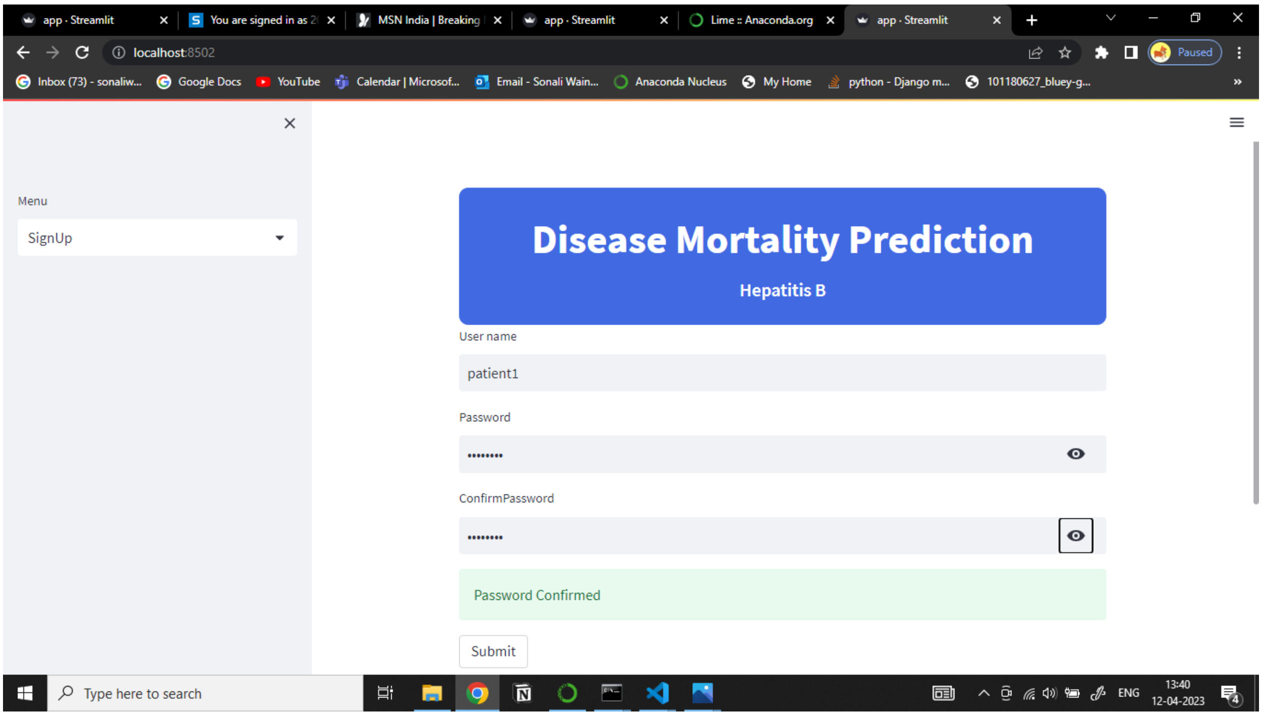


Figure 8.2: SignUp Page

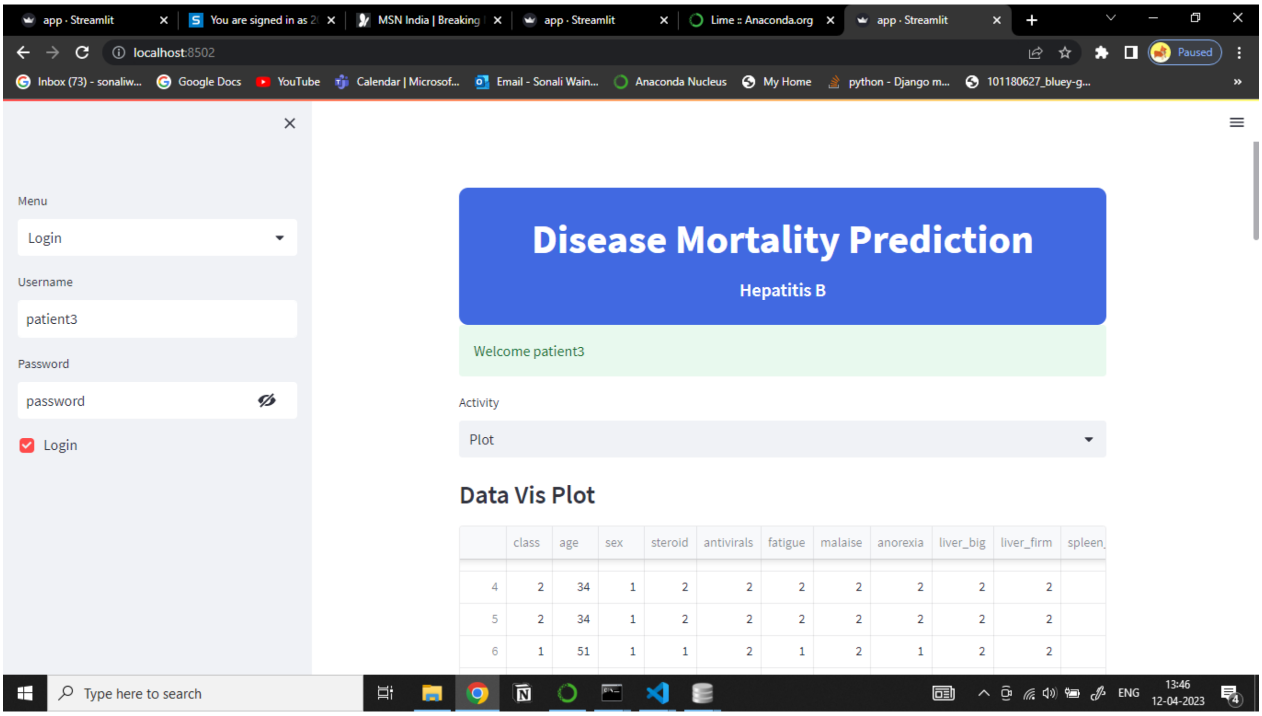


Figure 8.3: Login Page and Home Screen

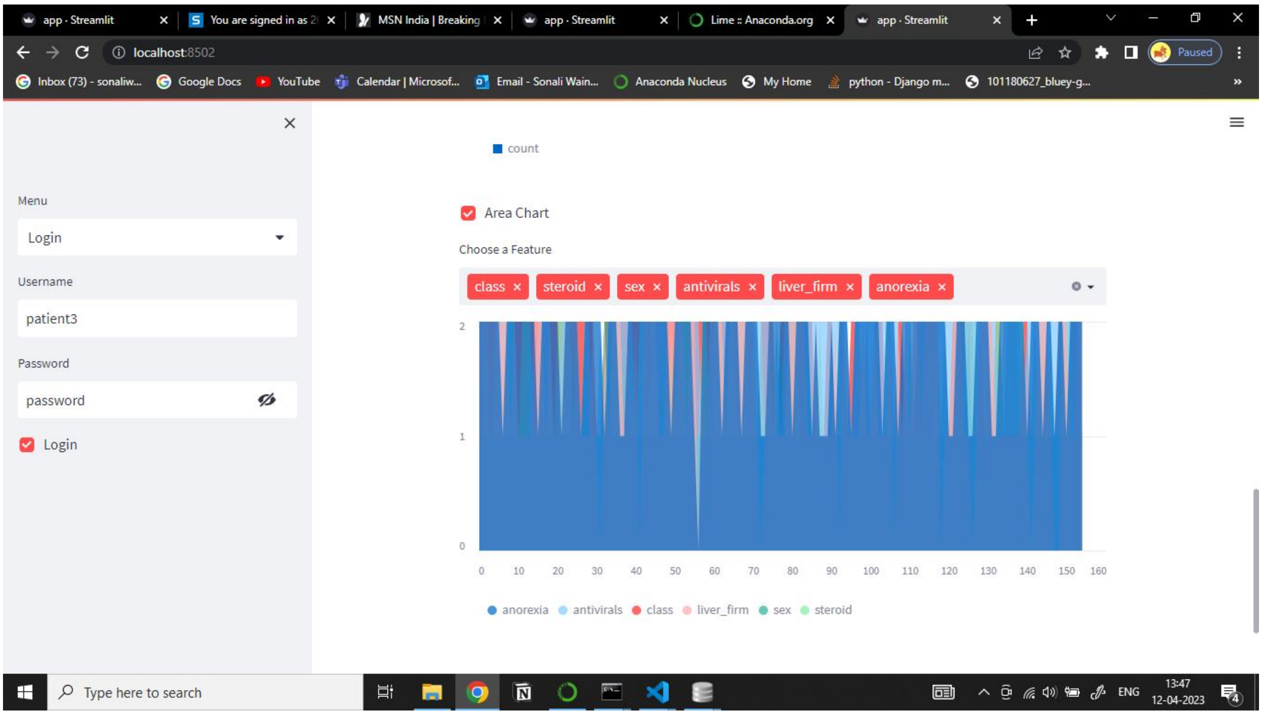


Figure 8.4: Area Chart of the Dataset

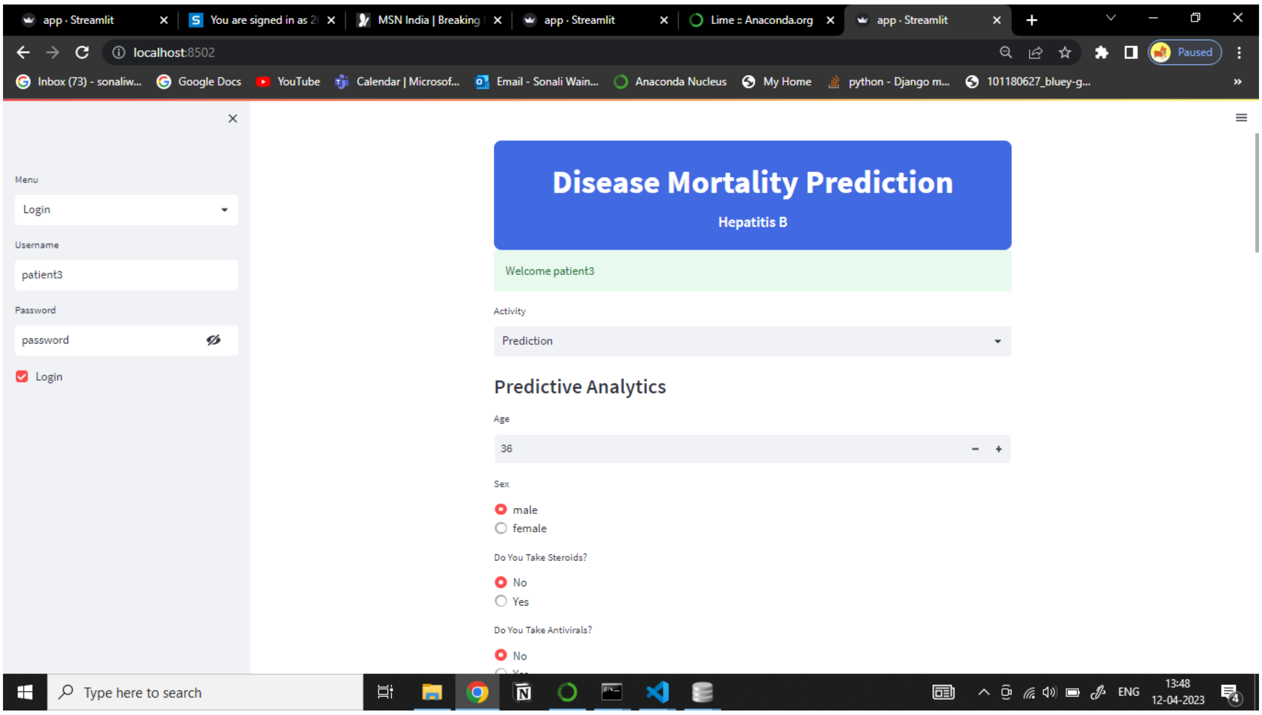


Figure 8.5: Predictive Analysis

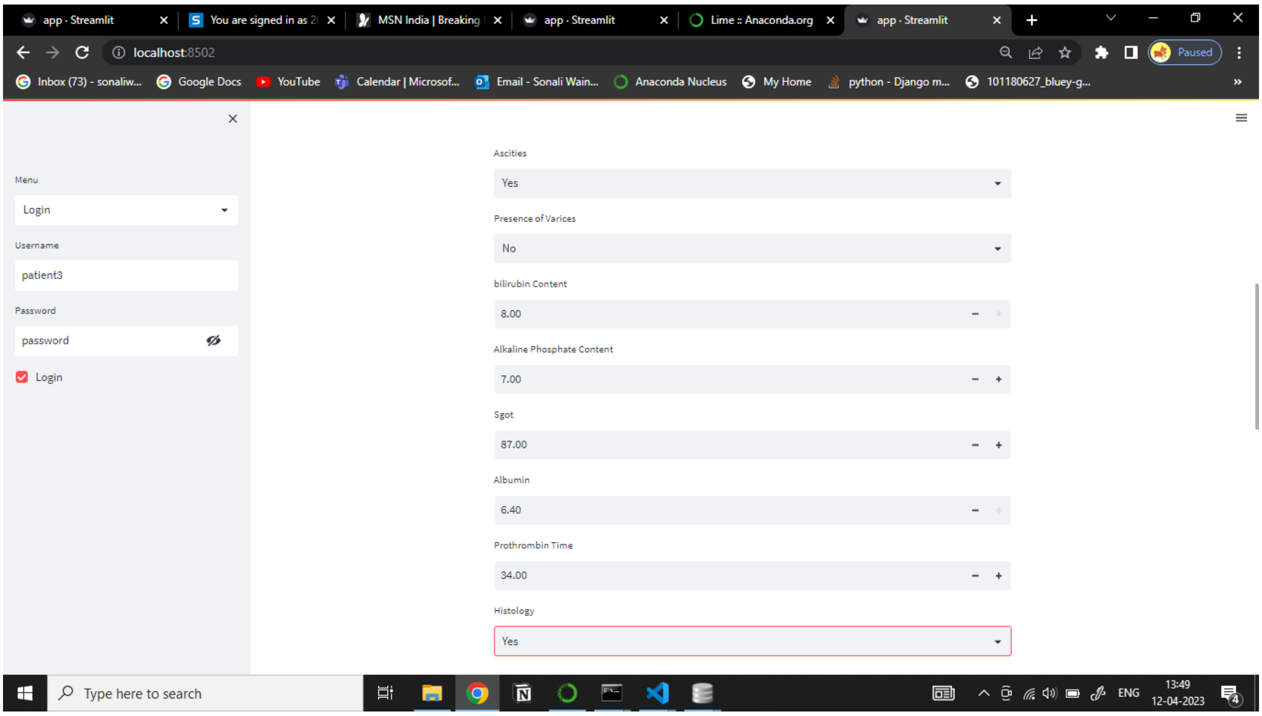


Figure 8.5: Data Input

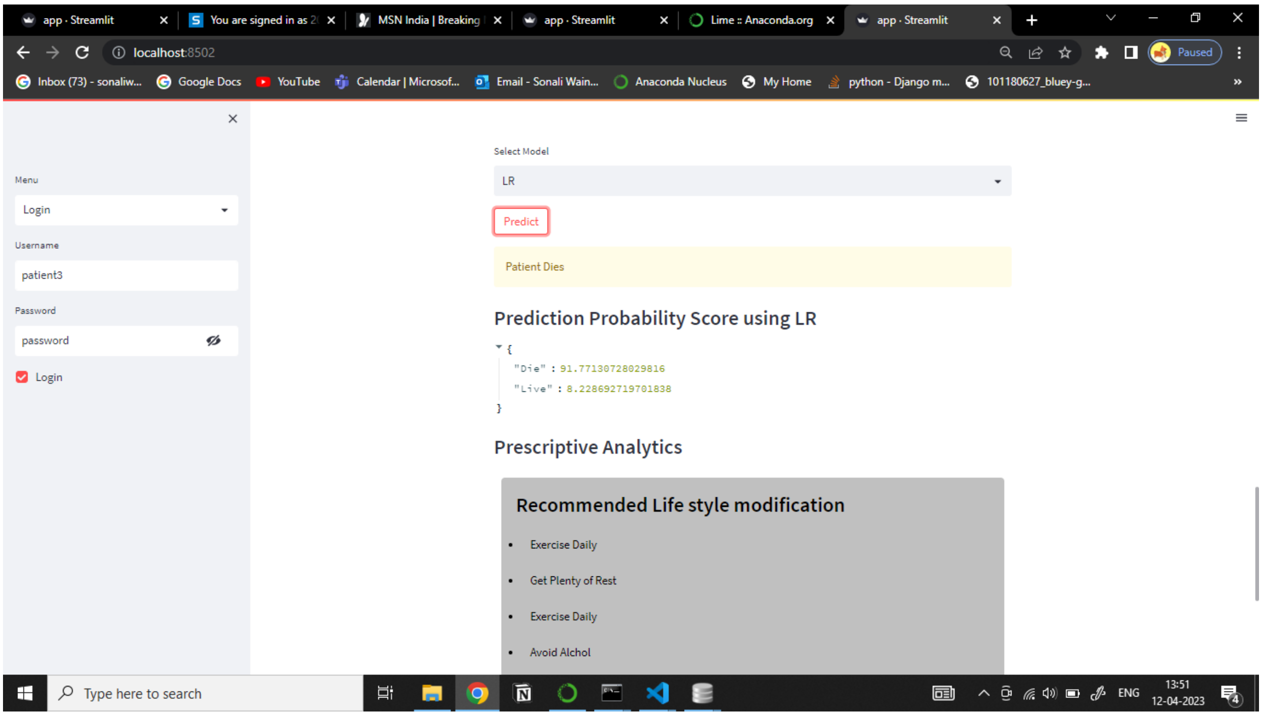


Figure 8.6: Prediction Probability Score using LR

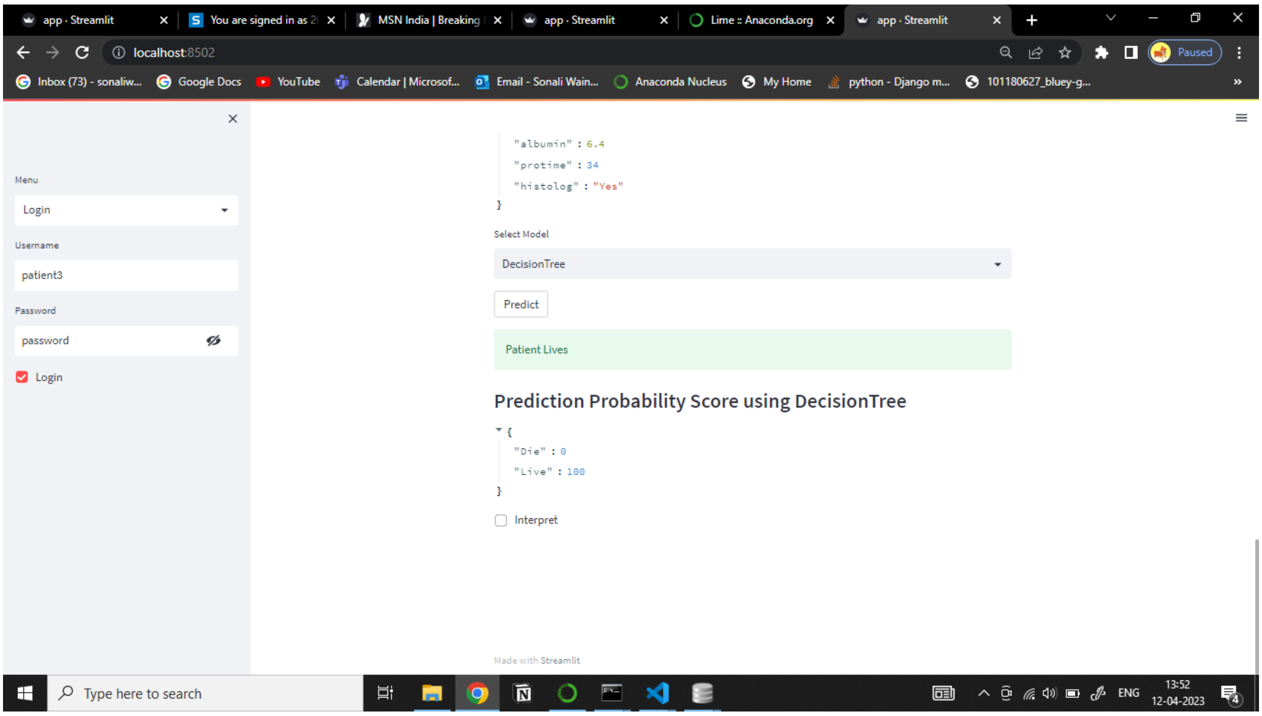


Figure 8.7: Prediction Probability Score using Decision

Chapter 9
Result and Discussion

- Suggestion in Review-1 is improve GUI.
- Suggestion in Review-1 is add Validations.

Chapter 10
Conclusion

In conclusion, the proposed hepatitis B mortality prediction system using machine learning can be a useful tool for healthcare providers in identifying patients at high risk of mortality due to hepatitis B. The system utilizes various features such as patient demographics, clinical factors, and laboratory results to train a predictive model that can accurately predict the likelihood of mortality. Additionally, the system provides decision support by stratifying patients into risk categories and providing appropriate interventions based on their risk level.

The functional requirements of the system include data collection, preprocessing, model training, prediction, risk stratification, and decision support. These requirements ensure that the system is able to accurately predict mortality risk and provide useful information to healthcare providers.

Overall, the proposed hepatitis B mortality prediction system has the potential to improve patient outcomes and reduce healthcare costs by identifying high-risk patients early and providing appropriate interventions. However, further research and validation are needed to ensure the accuracy and reliability of the system.

References

[1] Yang et al. (2019) "Comparison of Machine Learning Algorithms for Predicting Mortality in Patients with Hepatitis B"

[2] Lie et al. (2020)"Comparison of Machine Learning Algorithms for Predicting Mortality in Patients with Chronic Hepatitis B"

[3] Shao et. Al (2019)"Prediction of Hepatitis B-Related Mortality using a Decision Tree-Based Approach"