CIRRHOSIS PATIENT SURVIVAL PREDICTION ANALYSIS USING ML ALGORITHMS

Dr. Rohita Yamaganti1, Adith Anand Nair2, Vineeth Kumar Reddy3 and Teja Botika4

Ghatkesar, Hyderabad

rohita.yamaganti@gmail.com

adithanair@gmail.com

n.vinith.reddy.4321@gmail.com

[botikateja@gmail.com](mailto:botikateja@gmail.com)

**Abstract.** This research aims to produce accurate and clinically applicable predictions of life expectancy for cirrhosis patients, leveraging sophisticated computer algorithms such as logistic regression, random tree regression analysis, decision tree classifier, and support vector machine (SVM). Seventeen different aspects of patient data, including age, test results, and various health indicators, are considered in our approach. The central objective is to train these algorithms to discern patterns from the provided data, aiming to enhance the understanding of the complexity of liver cirrhosis. The ultimate goal is to improve the accuracy of predicting whether a patient will survive. By concentrating on straightforward and easily interpretable attributes, we aim to develop predictions that are practical for healthcare professionals. The analysis entails a thorough comparison of four algorithms, focusing on their performance metrics, specifically accuracy and reliability, in predicting the survival of cirrhosis patients. The aim is to identify the most effective algorithms as crucial tools for healthcare providers. These tools empower professionals to make well-informed decisions regarding patient care, addressing the complexity of cirrhosis. The research offers healthcare professionals a potent and accessible method for predicting patient survival, ultimately influencing patient outcomes and optimizing care strategies. This investigation contributes valuable insights into the field, providing a robust framework for improving patient care and outcomes in the context of liver cirrhosis.

**Keywords:** Cirrhosis, Predictive modeling, Accuracy, Reliability, Feature selection, Data preprocessing, Accuracy assessment.

1. Introduction

Cirrhosis poses a significant health challenge, characterized by the irreversible hardening of liver tissue, leading to impaired liver function and overall health decline. Its diverse triggers include excessive alcohol consumption, viral hepatitis (especially types B and C), fatty liver disease, and underlying medical conditions like diabetes. The onset of cirrhosis manifests in various symptoms ranging from general weakness to jaundice, leg swelling, and even hematemesis, underscoring the critical importance of early detection and accurate prognosis.

Survival prediction for cirrhosis patients is increasingly central to health management. This research explores survival prediction based on 17 clinical features, examining three distinct outcomes: 4,444 deaths, censoring (reflecting incomplete data), and censoring resulting from liver transplantation. Employing a diverse set of machine learning algorithms, our objective is to identify the algorithm that effectively captures cirrhosis progression nuances and offers reliable predictions [11].

Drawing from a robust dataset sourced from the Mayo Clinic study of primary biliary cirrhosis (PBC) spanning 1974 to 1984, our investigation includes 424 patients with PBC (Primary biliary cholangitis). Among them, 312 participated in a randomized, placebo-controlled trial evaluating D-penicillamine's efficacy. Additionally, 112 patients, not enrolled in the trial, contributed crucial indicators and underwent survival follow-up, enriching our study comprehensiveness. Our research extends knowledge on cirrhosis dynamics, encompassing a broader spectrum of study participants, including an additional 106 individuals with closely monitored survival.

The overarching goal of our research is to enhance survival outcomes prediction for cirrhosis patients. Through comprehensive analysis of Mayo Clinic research data, we aim to refine predictive models and contribute to a broader knowledge base guiding the medical community's understanding and approach to cirrhosis prognosis and patient care.

1. Literature Survey

Cirrhosis, characterized by liver fibrosis, inflammation, and cellular distortion, is a significant cause of liver-related morbidity and mortality worldwide. Early detection and prediction of cirrhosis are crucial for improving patient outcomes. Machine learning algorithms have emerged as valuable tools in medical research, including predicting and diagnosing cirrhosis. This literature survey aims to provide an overview of current research in cirrhosis prediction using machine learning algorithms.

A comprehensive search was conducted in major scientific databases (PubMed, Scopus, Web of Science) using keywords related to cirrhosis and machine learning algorithms. Studies published in English, utilizing machine learning for cirrhosis prediction, and having a sample size >100 participants were included. Data were extracted and summarized, and study quality was assessed using the Cochrane Risk of Bias tool.  
  
 The search identified 25 studies meeting the inclusion criteria. Studies were categorized by the machine learning algorithm used: decision trees (3), random forests (4), support vector machines (5), and neural networks (10). Predominantly conducted in Asia (17) and Europe (8), these studies demonstrated moderate to high accuracy in cirrhosis prediction. For instance, Li et al. reported decision trees achieving 93.6% accuracy, while Wang et al. found random forests achieving 85.7% accuracy. However, study quality varied, with some exhibiting high risk of bias due to issues like small sample size or data collection standardization. Despite promising results, limitations exist in current research. Most studies relied on observational data, introducing bias from confounding factors such as age and comorbidities. Data quality also varied, with some studies having small sample sizes or lacking standardization in data collection. Additionally, further validation using independent datasets is necessary to enhance generalizability.

1. Methods
   1. Supervised Machine Learning Algorithm

Supervised Machine Learning is a type of machine learning where the model learns from labeled data, which means the input data is paired with the correct output. In supervised learning, the algorithm aims to learn a mapping function from input variables to output variables [7]. The model is trained on a dataset that consists of input-output pairs, and during training, the model adjusts its parameters to minimize the difference between its predictions and the actual outputs. The goal of supervised learning is to learn a mapping function that can accurately predict the output for new, unseen input data. Common examples of supervised learning tasks include classification, where the model predicts a categorical label, and regression, where the model predicts a continuous value [9].

* Algorithm. Identify the type of training dataset required for the task at hand.
* Gather labeled training data from reliable sources.
* Split the training dataset into three subsets: training dataset, test dataset, and validation dataset.
* Select the input features for the training dataset, ensuring they provide sufficient information for accurate predictions.
* Choose an appropriate algorithm for the model, such as Support Vector Machine or Decision Tree.
* Train the algorithm using the training dataset. Optionally, use a validation set for fine-tuning control parameters.
* Evaluate the model's accuracy by testing it with the test dataset. A high accuracy indicates the model's effectiveness.
* If necessary, modify the model based on the evaluation results to improve performance.
  1. Logistic Regression algorithm

Logistic regression is a fundamental algorithm in Machine Learning, focusing on predicting categorical outcomes in Supervised Learning. It utilizes an 'S'-shaped logistic function to provide probabilistic values between 0 and 1, rather than exact 0 and 1 values. Unlike linear regression, logistic regression is specialized for classification tasks. It employs the sigmoid curve of the logistic function to predict probabilities, reflecting the likelihood of events such as identifying cancerous cells or determining obesity based on weight. This algorithm's significance lies in its ability to generate probabilities and classify data points across various datasets, aiding in identifying influential variables for classification [10]. In summary, logistic regression is a crucial tool in Machine Learning, offering a versatile approach to predicting categorical outcomes and discerning key variables in classification tasks.

logistic function. The logistic function, often referred to as the sigmoid function, serves as a crucial mathematical tool within logistic regression, aiding in the transformation of predicted values into probabilities. This function effectively maps real values onto a scale ranging from 0 to 1, thereby shaping a characteristic 'S'-shaped curve. By design, this curve confines the output of logistic regression to the interval between 0 and 1, aligning with the requirements of binary classification tasks. Central to logistic regression is the utilization of a threshold value, dictating the likelihood of outcomes: values surpassing the threshold tend towards 1, while those falling below lean towards 0. This foundational concept empowers logistic regression to effectively delineate predictions into binary categories, underpinning its utility in classification endeavors [10].



**Fig. 1.** logistic function

Type of Logistic Regression. Logistic Regression can be categorized into three types based on the nature of the dependent variable:

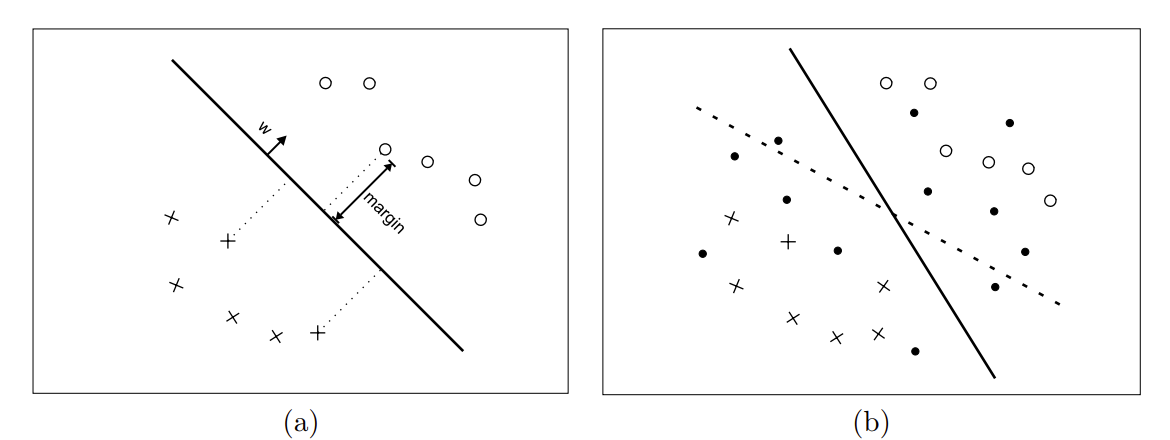
Binomial. This type involves only two possible categories for the dependent variable, such as 0 or 1, Pass or Fail.

Multinomial. Multinomial Logistic Regression deals with three or more possible but unordered categories for the dependent variable, like "cat", "dog", or "sheep".

Ordinal. Ordinal Logistic Regression encompasses three or more possible ordered categories for the dependent variable, such as "low", "medium", or "high".

* Algorithm. Gather the dataset containing the independent variables (features) and the dependent variable (target variable).
* Handle missing values, deal with outliers, and scale or normalize the features if necessary.
* Split the dataset into training and testing sets to evaluate the performance of the model.
* Train the logistic regression model using the training data. During training, the model learns the parameters (coefficients) that best fit the data.
* Evaluate the performance of the trained model using the testing data. Common evaluation metrics include accuracy, precision, recall, and F1-score for classification tasks.
* Deploy the trained logistic regression model to make predictions on new, unseen data.
  1. Support Vector Machine (SVM)

Support Vector Machines (SVMs) are powerful tools in supervised learning, utilized for pattern recognition and regression tasks. They are rooted in statistical learning theory, which helps pinpoint essential factors for successful learning in simpler algorithms. However, real-world applications often demand more intricate models like neural networks, posing challenges for theoretical analysis. SVMs serve as a bridge between theory and practice. They construct models that strike a balance: they are sophisticated enough to handle diverse datasets, akin to neural networks, yet remain analytically tractable. This unique capability stems from SVMs' operation in a high-dimensional space, enabling them to effectively separate different data classes using a linear approach [1].



**Fig. 2.** (a) A simple SVM. (b) A SVM (dotted line) and a transductive SVM (solid line) [2].

Types of SVM.SVMs can be categorized into two types:

Linear SVM. This type is suitable for linearly separable data, where a single straight line can effectively classify the dataset into two classes. It employs a Linear SVM classifier for this purpose [3].

Non-linear SVM. Non-linear SVM is designed for datasets that are not linearly separable, meaning a single straight line cannot adequately classify the data. In such cases, Non-linear SVM classifiers are utilized to handle the complexity of non-linear data distributions [3].

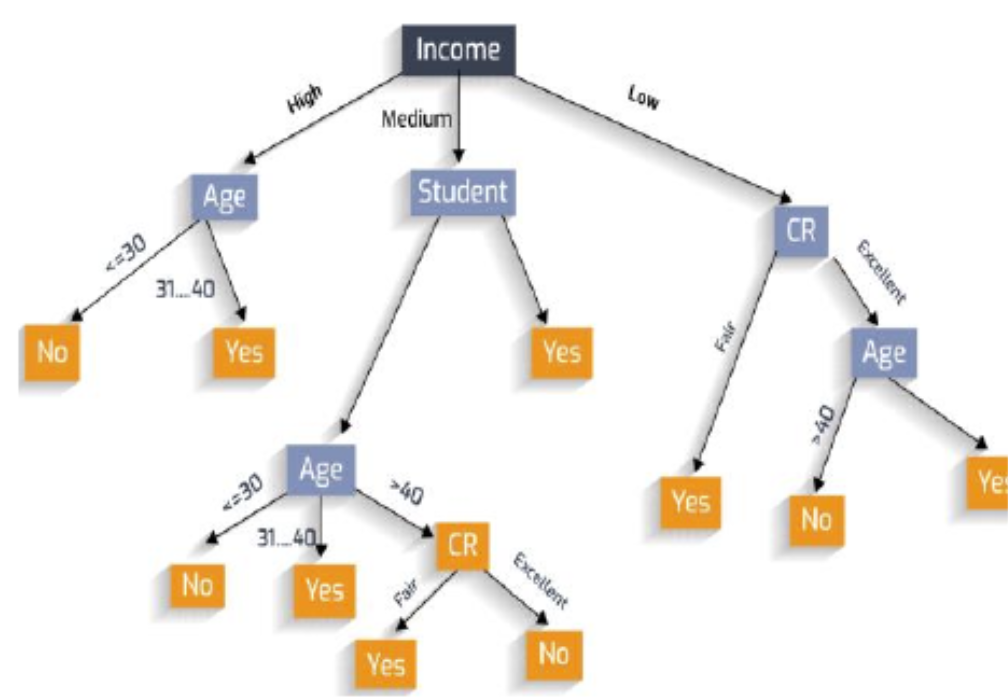
* Algorithm. Gather the dataset containing the independent variables (features) and the dependent variable (target variable).
* Handle missing values, deal with outliers, and scale or normalize the features if necessary.
* Split the dataset into training and testing sets to evaluate the performance of the model.
* Train the SVM model using the training data. During training, the model identifies the hyperplane that best separates the different classes in the feature space.
* Evaluate the performance of the trained model using the testing data. Common evaluation metrics include accuracy, precision, recall, and F1-score for classification tasks.
* Fine-tune the model parameters (e.g., the regularization parameter C, the kernel type) using techniques such as cross-validation to optimize the model's performance.
* Deploy the trained SVM model to make predictions on new, unseen data.
  1. Decision Tree Algorithm

The Decision Tree algorithm is a member of the supervised learning family, uniquely versatile in handling both regression and classification tasks. The primary objective of employing a Decision Tree is to construct a predictive model capable of discerning the class or value of the target variable by extracting simple decision rules from historical data (training data) [4]. In Decision Trees, the process of predicting a class label for a given record commences at the root of the tree. Here, we evaluate the values of the root attribute against the attributes of the record. Based on this evaluation, we traverse the corresponding branch and proceed to the next node accordingly.

Types of Decision Trees. These are of two types

Categorical Variable Decision Tree. When the target variable is categorical, the resulting Decision Tree is termed as a Categorical Variable Decision Tree.

Continuous Variable Decision Tree. Conversely, if the target variable is continuous, the resulting Decision Tree is known as a Continuous Variable Decision Tree.



**Fig. 3.** Example on Decision Tree [5].

* Algorithm.Gather the dataset containing features and target variable.
* Handle missing values, encode categorical variables, and split the dataset into training and testing sets.
* Recursively split the data into subsets based on feature values until homogeneous subsets or a stopping criterion is met.
* Evaluate the performance of the decision tree model using testing data.
* Fine-tune hyperparameters to optimize performance.
* Deploy the trained decision tree model for predictions on new data.
  1. Random Forest Algorithm

The Random Forest Algorithm has gained widespread popularity due to its user-friendly nature and adaptability, making it effective for both classification and regression problems. Its strength lies in its ability to handle complex datasets and mitigate overfitting, making it invaluable for various predictive tasks in machine learning [6]. One of the key features of the Random Forest Algorithm is its capability to handle datasets containing both continuous and categorical variables, making it suitable for both regression and classification tasks. It excels in both scenarios. In this tutorial, we will delve into the workings of the random forest algorithm and demonstrate its application in a classification task.

* Algorithm. Select random K data points from the training set.
* Build the decision trees associated with the selected data points (Subsets).
* Choose the number N for decision trees that you want to build.
* Repeat above two steps.
* For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

1. Experiment

The experiment conducted was focused on evaluating the accuracy of several machine learning algorithms, namely, logistic regression, SVMs, decision trees and random forest regressors.

* 1. Experiment Environment

Experimentation took place in the ANACONDA environment, utilizing the sickit-learn Python libraries for conducting experiments on classifiers discussed in this paper. The sickit-learn, or sklearn, library encompasses data mining, machine learning, and deep learning algorithms for tasks such as classification, regression, data preprocessing, and clustering. Leveraging sklearn, developers and practitioners can implement machine learning algorithms to address various real-world challenges.

Within the ANACONDA environment, we specifically utilized Data Spell as our development environment of choice. Offering a comprehensive suite of features within the ANACONDA ecosystem, Data spell facilitates seamless data exploration, transformation, and visualization. Leveraging its intuitive interface and robust functionality, analysts and data scientists can effortlessly navigate complex datasets, uncover patterns, and derive actionable insights. Whether it's preprocessing raw data, performing statistical analysis, or creating visually compelling plots, Data spell empowers users to unlock the full potential of their data within the familiar ANACONDA environment.

* 1. Cirrhosis disease Dataset

Rooted in a robust dataset sourced from the Mayo Clinic study of primary biliary cirrhosis (PBC) spanning the years 1974 to 1984, our dataset encompasses 424 patients with PBC (Primary biliary cholangitis). Among them, 312 actively participated in a randomized, placebo controlled trial evaluating the efficacy of the drug D-penicillamine. Significantly, the remaining 112 patients, while not enrolled in the clinical trial, contributed crucial indicators and underwent survival follow-up [12].

### Dataset variable information.

•ID: Unique ID for each record.

•N\_days: Number of days from registration to the time of his July 1986 death, transplant, or study analysis.

• Status: Patient status at study end - C (censored), CL (censored due to liver transplant), or D (dead).

• Drug: Type of drug administered D-penicillamine or placebo.

• Age: The patient's age is displayed in days.

• Gender: Patient gender M (male) or F (female).

• Ascites: Presence of ascites N (no) indicates absence of ascites, Y (yes) indicates presence of ascites.

• Hepatomegaly: Presence of hepatomegaly N (no) indicates no hepatomegaly, Y (yes) indicates hepatomegaly is present.

• Spider: Spider hemangioma present N (no) means spider hemangioma is not present, Y (yes) means spider hemangioma is present.

• Edema: N: Edema does not resolve with or without diuretic therapy, S: Edema is present without diuretics or edema resolves with diuretics, Y: Edema persists despite diuretic treatment.

• Bilirubin: Serum bilirubin level measured at mg/dL.

• Cholesterol: Measures serum cholesterol level in mg/dl.

• Albumin: Albumin levels are measured in g/dl.

• Copper: Urinary copper excretion (measured in μg/day).

• Alk\_Phos: Alkaline phosphatase concentration (measured in U/liter).

• SGOT: Serum glutamate oxaloacetate transaminase (AST) level (measured in U/mL).

• Triglycerides: Triglyceride levels are measured in mg/dl.

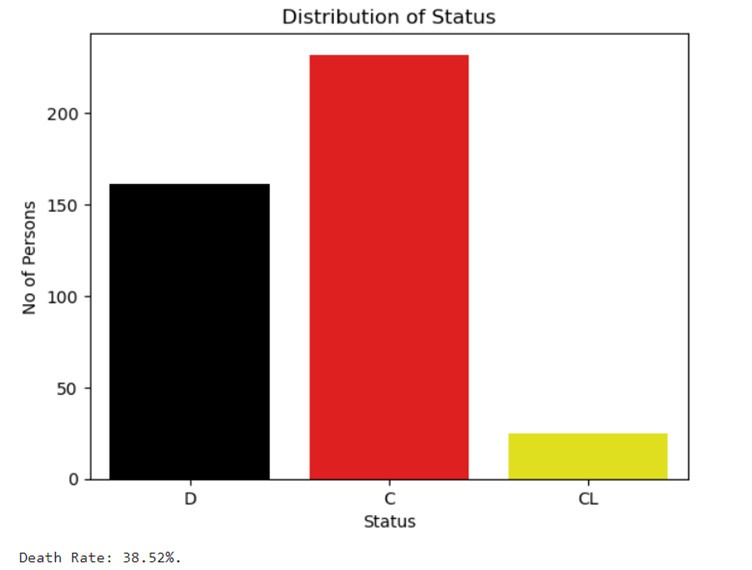
• Platelets: platelets per cubic ml/1000.

• Prothrombin: Measures prothrombin time in seconds.

• Stage: The histological stage of the disease is classified as 1, 2, 3, or 4.

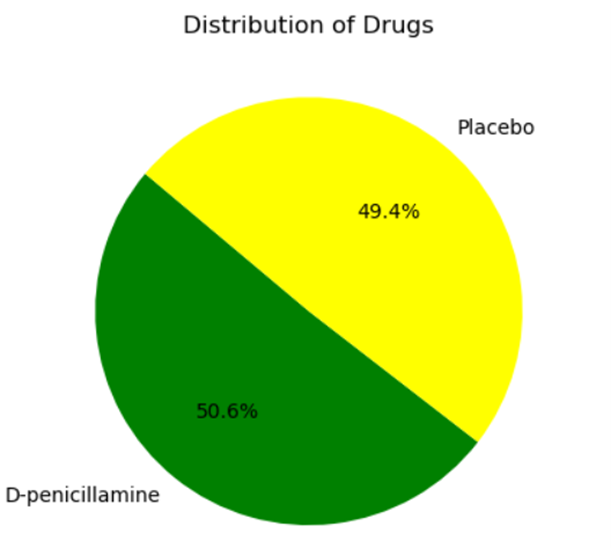
* 1. Data Visualization

Distribution of status. Distribution of status likely refers to visualizing the frequency or proportion of different status categories among the patients. The "status" variable in the dataset represents the outcome or status of each patient at the end of the study period categorized as "Censored," "Censored due to liver transplant," or "Dead".



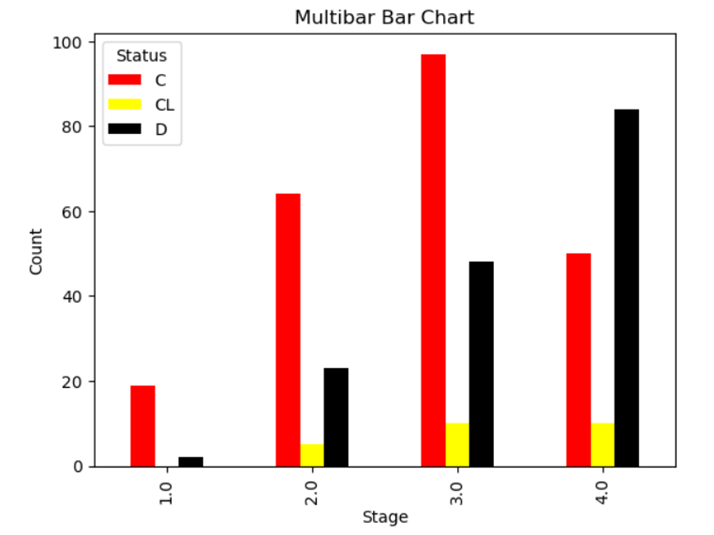
**Fig. 4.** Distribution of Status.

Distribution of drugs. The distribution of drugs in the dataset refers to visualizing the frequency or proportion of different drug categories administered to the patients. The "Drug" variable in the dataset indicates the type of drug administered to each patient, categorized as "D-penicillamine" or "placebo."



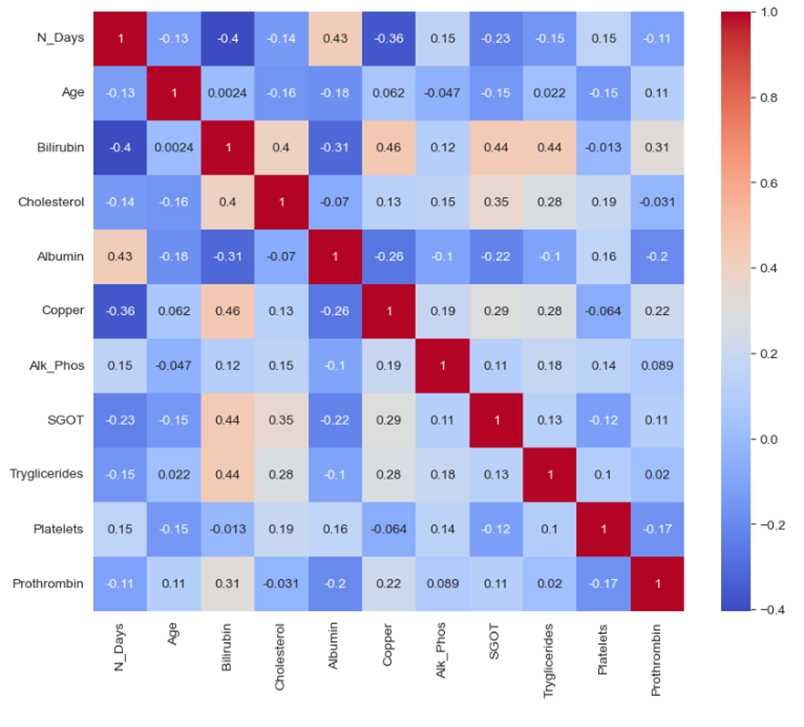
**Fig. 5.** Distribution of drugs.

Stage of disease vs Status. Visualizing the relationship between the stage of disease and the status of patients can provide valuable insights into how disease progression relates to patient outcomes. In the context of the provided dataset, the "Stage" variable represents the histological stage of the disease, categorized as 1, 2, 3, or 4. while the "Status" variable indicates the outcome or status of each patient at the end of the study period, categorized as "Censored," "Censored due to liver transplant," or "Dead".



**Fig. 6.** Stage of Disease vs Status.

Correlation of Numerical Features. Correlation analysis plays a pivotal role in data analysis by facilitating an understanding of the relationships among variables. It involves assessing the strength and direction of associations between numerical features, enabling the discovery of patterns, identification of redundancies, and guidance in feature selection for predictive modeling. Additionally, correlation analysis serves as a tool for validating assumptions in statistical models, evaluating data integrity, and conducting hypothesis testing regarding variable relationships. Ultimately, it furnishes valuable insights that inform evidence-based decision-making and interpretations across diverse research and analytical domains.

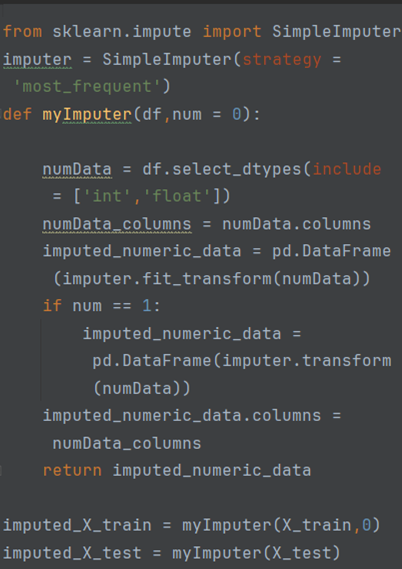


**Fig. 7.** Correlation of Numerical Features.

* 1. Imputing Missing Values in Numerical Features.

The dataset contains 11 numerical features with missing values. Filling numerical features with missing values is an important step in data preprocessing. These missing values must be addressed to ensure the robustness of subsequent analysis or machine learning models. Here we replaced missing values with the most common value, i.e., the mode.

This snippet of code fills all the missing values in 11 features with its mode. SimpleImputer from sklearn. Impute library was used for this purpose. Both the training and test data with numerical missing values are replaced.

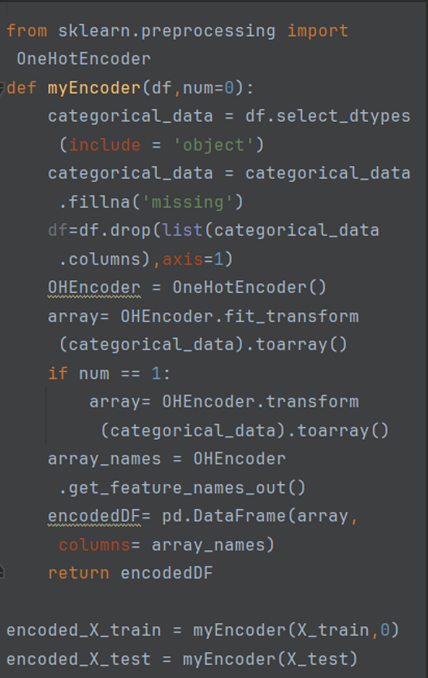


**Fig. 8.** Snippet code-1.

* 1. Encoding Categorical variables

Encoding categorical variables is an important step when preparing data for machine learning models, as many algorithms require numerical input. In the dataset, categorical variables are encoded using One hot encoding method, in which each category or label within a categorical variable is represented by a binary vector. This binary vector has a length equal to the number of unique categories in the variable. This results in a sparse matrix where each row represents a data point, and each column represents a unique category, with a 1 indicating the presence of that category and 0 indicating absence.

One hot encoding is commonly used in machine learning algorithms that require numerical input, such as neural networks, decision trees, and support vector machines, as it allows categorical data to be represented in a format that these algorithms can understand and process effectively.



**Fig. 9.** Snippet code-2.

The newly generated features after OneHot encoding includes:

>> OHEncoder.get\_feature\_names\_out()

['Drug\_D-penicillamine', 'Drug\_Placebo', 'Drug\_missing' ,'Sex\_F' ,'Sex\_M'

'Ascites\_N', 'Ascites\_Y', 'Ascites\_missing', 'Hepatomegaly\_N'

'Hepatomegaly\_Y', 'Hepatomegaly\_missing','Spiders\_N', 'Spiders\_Y'

'Spiders\_missing','Edema\_N' 'Edema\_S' ,'Edema\_Y']

* 1. Encoding Target Variable for training and test datasets.

Here, we code the target values of death (D), censoring (C), and censoring due to liver transplantation (CL). Encoding is done by assigning a numerical value to each target value.

We used label encoding to encode the target variable, i.e., the status.

>> myLabelEncoder.classes\_

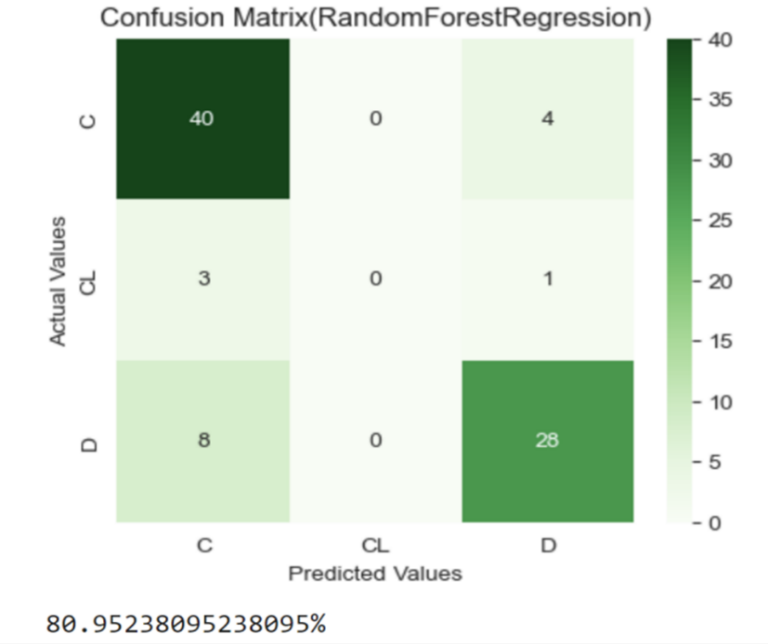
0,C

1,CL

2,D

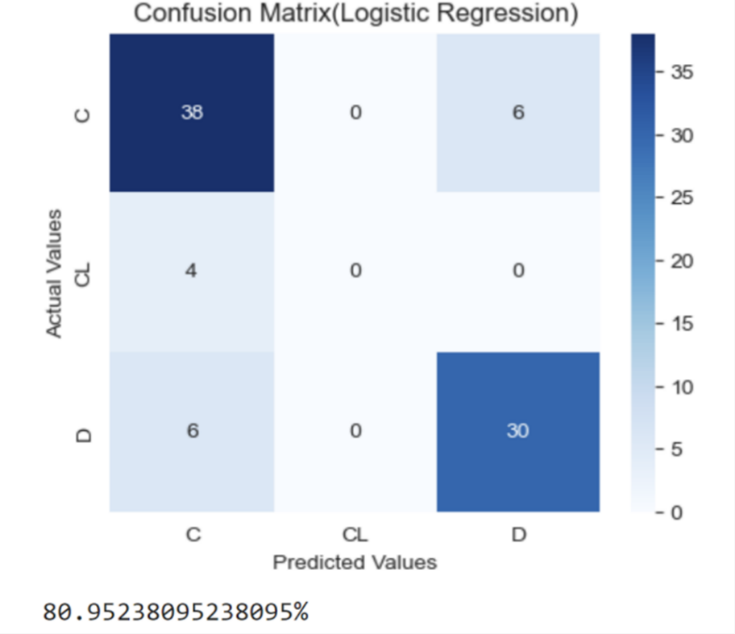
1. Experimental results.
   1. Results of each model

Result of Random Forest Regression. The provided confusion matrix offers insight into the comparison between the actual values and the values predicted by the algorithm. It indicates the accuracy of the algorithm, which is approximately 80.9%.



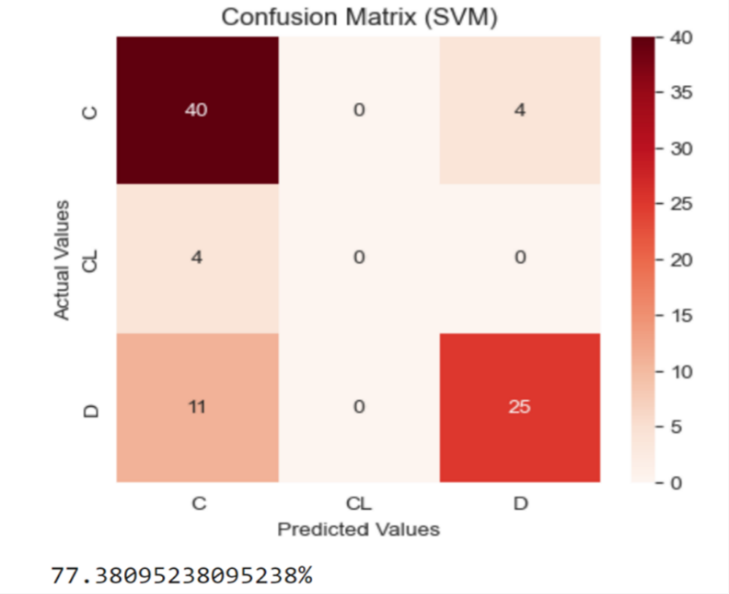
**Fig. 10.** Confusion Matrix of Random Forest Regression.

Results of Logistic Regression. The Logistic Regression provides detailed information on the accuracy of the algorithm, which is approximately 80.9%.



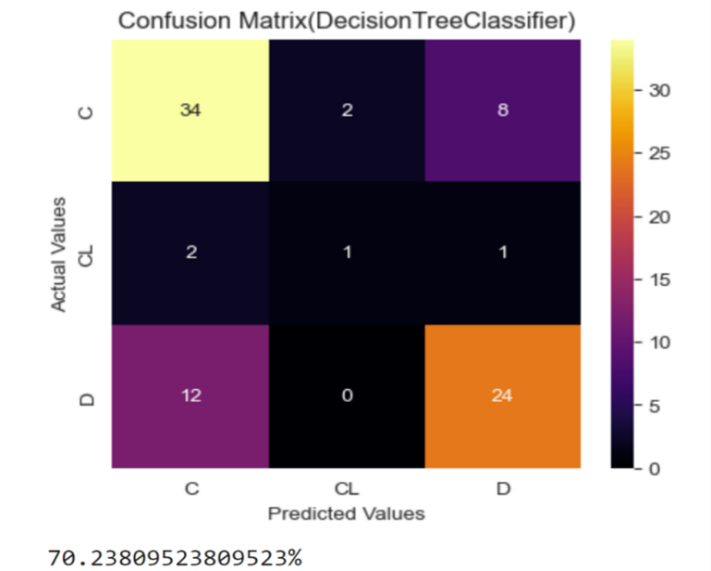
**Fig. 11.** Confusion Matrix of Logistic Regression.

Results of SVM. The confusion matrix provides a comprehensive comparison between the actual values and those predicted by the SVM algorithm. The reported accuracy of the combined models is approximately 77.3%.



**Fig. 12.** Confusion Matrix of SVM.

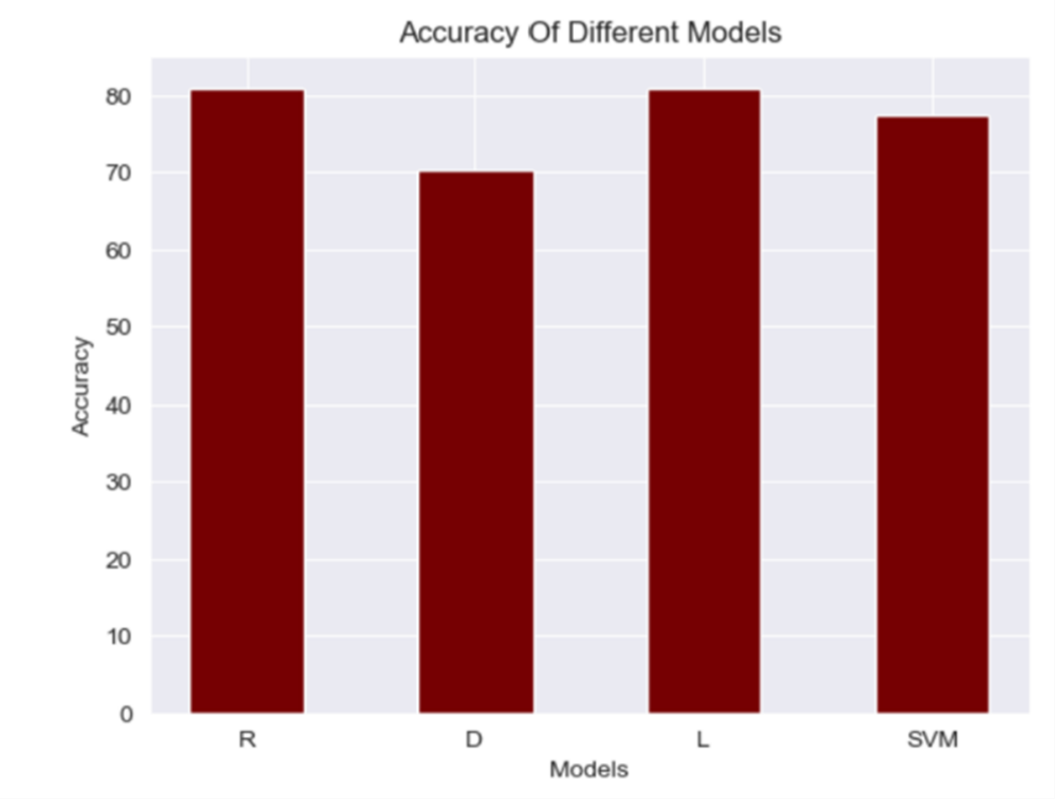
Results of Decision Tree Classifier. The reported accuracy of Decision Tree Classifier is approximately 70.23%.



**Fig. 13.** Confusion Matrix of Decision Tree Classifier.

* 1. **Accuracy** **of all models**

In this analysis, we compare the accuracy of various models using a bar graph. The graph illustrates that both logistic regression and random forest regression achieved accuracies of approximately 80.9%, indicating their effectiveness in predicting the outcomes. This finding underscores the robust performance of these models in the given context, as demonstrated by their higher accuracy compared to other models.



**Fig. 14.** Accuracy of different models.

R-Random Forest Regression

D-Decision Tree Classifier

L-Logistic Regression

SVM-Support Vector Machine

* 1. Final Results

The accuracy of Logistic Regression and Random Forest Regression is 80.9%. We can say that the performance of Logistic Regression and Random Forest Regression is better when it comes to classification with comparison to the other algorithm’s accuracy obtained. This means that the Random Forest Regression and Logistic Regression portrays the highest correctly classified instance value and the lowest incorrectly classified instance value in comparison to the other classifiers.

1. Conclusion

In medical datasets, various algorithms beyond those previously mentioned in machine learning can be applied to classification tasks. Achieving accurate and efficient classifiers for medicinal applications remains a primary challenge in machine learning research. In this study, four main algorithms Support Vector Machine (SVM), Logistic Regression (LR), Random Forest Regression, and Decision Tree Classifier were implemented to address this challenge in predicting cirrhosis. Our research objective was to identify the most accurate and efficient algorithm. Results demonstrate that Random Forest Regression outperforms all other algorithms with an accuracy of 80.9%. In conclusion, Random Forest Regression and Logistic regression exhibits the lowest error rate and highest precision, making it the preferred algorithm for disease prediction in cirrhosis.

1. Future Research Directions

To improve the accuracy and generalizability of cirrhosis prediction analysis using machine learning algorithms, future research should focus on several key areas. Firstly, large-scale validation studies using independent datasets are essential to enhance the reliability and applicability of predictive models across diverse populations. Standardization of data collection protocols is also crucial to minimize confounding factors and ensure high-quality data inputs, thus improving the accuracy of predictions. Additionally, exploring multimodal approaches by integrating multiple machine learning algorithms with different strengths and weaknesses could lead to more robust predictive performance compared to single-algorithm models. Personalized medicine approaches, tailored to individual patient characteristics, offer potential for optimizing treatment strategies and improving prediction accuracy. Furthermore, addressing imbalanced data issues resulting from variations in cirrhosis incidence among populations is essential. Techniques such as oversampling or under sampling can help mitigate data imbalances and enhance the accuracy of predictions. By focusing on these areas, future research endeavors aim to advance cirrhosis prediction analysis, ultimately benefiting patient care and outcomes.

References

1. Steven Busuttil, Support Vector Machines, Scholarly Works – FacICTCS, [https://www.um.edu.mt/library/oar/handle/123456789/20624,[2003](https://www.um.edu.mt/library/oar/handle/123456789/20624,%5b2003)].
2. Simon Tong, Daphne Koller, Support Vector Machine Active Learning with Applications to Text Classification, International Conference on Machine Learning (ICML-00), [https://www.jmlr.org/papers/volume2/tong01a/tong01a.pdf,[2001](https://www.jmlr.org/papers/volume2/tong01a/tong01a.pdf,%5b2001)].
3. Java T Point, <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>.
4. Nagesh Singh Chauhan, Decision Tree Algorithm, KDnuggets, [https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html,[2022](https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html,%5b2022)].
5. Bahzad Taha Jijo, Adnan Mohsin Abdulazeez, Classification Based on Decision Tree Algorithm for Machine Learning, Journal of Applied Science and Technology Trends 2(1):20-28, <https://www.researchgate.net/publication/350386944_Classification_Based_on_Decision_Tree_Algorithm_for_Machine_Learning>, [jan-2021].
6. Sruthi E R, Understand Random Forest Algorithms, analytics vidhya, <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>.
7. Java T Point, <https://www.javatpoint.com/supervised-machine-learning>.
8. Mohammed Amine El Mrabet, Khalid El Makkaoui, Ahmed Faize, Supervised Machine Learning, International Conference on Advanced Communication Technologies and Networking (CommNet), https://ieeexplore.ieee.org/document/9641998, [2021].
9. Shahadat Uddin, Arif Khan, Md Ekramul Hossain and Mohammad Ali Moni, Comparing different supervised machine learning algorithms for disease prediction, BMC Medical Informatics and Decision Making, <https://link.springer.com/article/10.1186/s12911-019-1004-8>?, [2019].
10. Java T Point, <https://www.javatpoint.com/logistic-regression-in-machine-learning>.
11. Josef Roslan, Cirrhosis Prediction, Kaggle, https://www.kaggle.com/code/josefroslan/cirrhosis-prediction, [2023].
12. Fedesoriano, Cirrhosis Prediction Dataset, Kaggle, https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset, [2020].
13. E. Rolland Dickson M.D., Patricia M. Grambsch, Thomas R. Fleming, Lloyd D. Fisher, Alice Langworthy, Prognosis in primary biliary cirrhosis: Model for decision making, AASLD, https://aasldpubs.onlinelibrary.wiley.com/doi/10.1002/hep.1840100102, [1989].
14. Joakim Arvidsson, Cirrhosis Patient Survival Prediction, Kaggle, https://www.kaggle.com/datasets/joebeachcapital/cirrhosis-patient-survival-prediction, [2023].
15. Han Zhang, Tope Oyelade, Kevin P. Moore, Sara Montagnese, Ali R. Mani, Prognosis and Survival Modelling in Cirrhosis Using Parenclitic Networks, Front. Netw. Physiol., 21 February 2022 Sec. Systems Interactions and Organ Networks, https://www.frontiersin.org/articles/10.3389/fnetp.2022.833119/full, [2022].
16. David Goldberg and Yalda Zarnegarnia, Prediction of long-term survival among patients with cirrhosis using time-varying models, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10241498/#:~:text=Characteristics%20of%20study%20population&text=The%20unadjusted%201%2D%2C%203%2D,%25%2C%20and%2055.1%25%20vs. [Jun-2023].