

# LITERATURE SURVEY

## INTRODUCTION:

Liver diseases avert the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises. Early prediction of liver disease using classification algorithms is an efficacious task that can help the doctors to diagnose the disease within a short duration of time. Discovering the existence of liver disease at an early stage is a complex task for the doctors. The main objective of this project is to analyze the parameters of various classification algorithms and compare their predictive accuracies to find out the best classifier for determining the liver disease.

## LITERATURE REVIEW:

### 1) Paper Name : Liver Disease Prediction System using Machine Learning Techniques

Authors: Rakshith D B, Mrigank Srivastava, Ashwani Kumar,  
Gururaj S P

Link: [https://github.com/IBM-EPBL/IBM-Project-21885-1659795062/blob/main/Project%20Design%20and%20Planning/Research\\_paper\\_1.pdf](https://github.com/IBM-EPBL/IBM-Project-21885-1659795062/blob/main/Project%20Design%20and%20Planning/Research_paper_1.pdf)

### Content:

In this paper we are going to discuss how to predict risk of liver disease for a person, based on the blood test report results of the user. In this paper, the risk of liver disease was predicted using various machine learning algorithms. The final output was predicted based on the most accurate machine learning algorithm. Based on the accurate model we designed a

system which asks a person to enter the details of his/her blood test report. Then the system uses the most accurate model which is trained to predict whether a person has risk of liver disease or not.

## **2) Paper Name: A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms**

**Authors:** A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain

**Link:** [https://github.com/IBM-EPBL/IBM-Project-21885-1659795062/blob/main/Project%20Design%20and%20Planning/Research\\_paper\\_2.pdf](https://github.com/IBM-EPBL/IBM-Project-21885-1659795062/blob/main/Project%20Design%20and%20Planning/Research_paper_2.pdf)

### **Content:**

Chronic Liver Disease is the leading cause of global death that impacts the massive quantity of humans around the world. This disease is caused by an assortment of elements that harm the liver. For example, obesity, an undiagnosed hepatitis infection, alcohol misuse. Which is responsible for abnormal nerve function, coughing up or vomiting blood, kidney failure, liver failure, jaundice, liver encephalopathy and there are many more. This disease diagnosis is very costly and complicated. Therefore, the goal of this work is to evaluate the performance of different Machine Learning algorithms in order to reduce the high cost of chronic liver disease diagnosis by prediction. In this work, we used six algorithms Logistic Regression, K Nearest Neighbors, Decision Tree, Support Vector Machine, Naïve Bayes, and Random Forest. The performance of different Classification techniques was evaluated on different measurement techniques such as accuracy, precision, recall, f-1 score, and specificity. We found the accuracy 75%, 74%, 69%, 64%, 62% and 53% for LR, RF, DT, SVM, KNN and NB. The analysis result shown the LR achieved the highest accuracy. Moreover, our present study mainly focused on the use of clinical

data for liver disease prediction and explore different ways of representing such data through our analysis.

**3)Paper Name: Liver Disease Prediction Using Machine Learning Algorithms: A Comparative Study**

**Authors:** Md. Fazle Rabbi, S.M. Mahedy Hasan, Arifa Islam Champa, Md. Asif Zaman, Md. Kamrul Hasan

**link:** <https://ieeexplore.ieee.org/document/9333528/>

**Content :**

In this paper, we have compared several ML methods, such as Logistic Regression (LR), Decision Tree (DT), Random Forest (RF) and Extra Trees (ET) for the prediction of liver disorders. Issues ignored by the previous researchers have been taken into consideration to improve the prediction accuracy. At the preprocessing steps, categorical values are encoded through label encoding. Then we have utilized the Pearson Correlation Coefficient to improve irrelevant features. After the removal of redundant features, over sampling is used to mitigate the imbalanced class distribution problem, and feature scaling is used to handle the outliers. After the completion of data preprocessing, LR, DT, RF, and ET classifiers are used for classifying liver disorders patients. For further improvement, the AdaBoost classifiers are used to increase the performance of each classification algorithm.

**4)Paper Name: Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques**

**Authors:** Satessh Ambesange, Vijayalaxmi A, Rashmi Uppin, Shruthi Patil, Vilaskumar Patil

**Link:** <https://ieeexplore.ieee.org/document/9500023/>

**Content:**

The machine-learning model is capable of predicting diseases, based on a data set, which is built in combination of key health parameters of a person with diseases and without diseases. For building models, an effective data set is needed, with proper representation of disease classifications. In this work, Indian Liver Patient Dataset hosted at [ics.uci.edu](http://ics.uci.edu) is used. Several classification machine-learning algorithms are able to classify liver diseases. Instead of selecting the algorithm, which gives better performance, the paper approaches how to tune the ML module for Random Forest algorithm in step-by-step ways. The Random Forest algorithm is to build a model, since it is trained on several samples of data obtained from splitting data so that the model is not tuned for accurate data. The main focus of the paper is to deeply analyze how models can be further tuned beyond one point of saturation due to an imbalanced data set. Various balancing techniques discussed and their impacts on performance are tabulated in later sections.