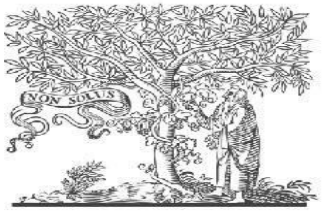
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 31st Mar 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 03

10.48047/IJIEMR/V12/ISSUE 03/93

Title **PREDICTION OF LIVER DISEASE USING LOGISTIC REGRESSION AND RANDOM FOREST ALGORITHM**

Volume 12, ISSUE 03, Pages: 640-644

Paper Authors
Mr. Vidyasagar, Keerthi G, Sowmya K, Salma M

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Prediction of Liver Disease Using Logistic Regression and Random Forest Algorithm

Mr. Vidyasagar V, Asst.prof, Department of Information Technology, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India
Keerthi G, Dept of Information Technology Andhra Loyola Institute of Engineering And Technology, Vijayawada, India
Sowmya K, Dept of Information Technology Andhra Loyola Institute of Engineering and Technology, Vijayawada, India
Salma M, Dept of Information Technology Andhra Loyola Institute of Engineering and Technology, Vijayawada, India

Abstract—It's crucial to predict liver disease in its early stages. In contrast to urban areas and major cities, where liver disease is more severe and currently more prevalent, rural areas have a low prevalence of the condition. Millions of people die each year as a result of liver disease. Early liver illness does not reveal any concerns. Patients' chances of survival can be improved by early detection of liver disease complications. In this project, we used a dataset of Indian patients with liver disease that included information on age, gender, total and direct bilirubin levels, alkaline phosphatase, alanine and aspartate aminotransferases, total proteins, albumin and globulin ratios, results, and 416 patients with liver disease and 167 patients without liver disease. The project seeks to forecast liver illness based on the user's blood test report results. The project's primary area of focus is machine learning, which covers data science and artificial intelligence. We estimate the risk of liver illness using machine learning methods. Random Forest and Logistic Regression are two machine learning methods used in this project. The research then utilizes the model's training data to predict whether or not a person has liver disease.

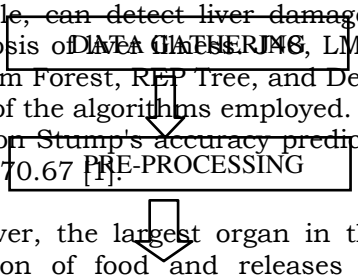
Keywords-Random Forest Algorithm, Logistic Regression Algorithm, Indian liver patient datasets

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Introduction

The liver is a vital organ in the human body. Changes in the color of the urine, stomach discomfort, jaundice (yellowing of the skin and eyes), nausea, back pain, vomiting, abdominal swelling, pale stools, exhaustion, fluid in an atypical cavity, and enlarged spleen and gall bladder are all signs of liver disease. Sometimes liver disease has no symptoms. Typical signs of liver illness can include right upper quadrant abdominal discomfort, vomiting, nausea, backpain, abdominal pain, exhaustion and weakness. Imaging and liver function tests, for example, can detect liver damage and aid in the prognosis of liver illness. J48, LMT, Random Tree, Random Forest, REP Tree, and Decision Stump are some of the algorithms employed.

Decision Stump's accuracy prediction was greater; it was 70.67 [1].



The liver, the largest organ in the body, aids in digestion of food and releases toxins from the body. One of the primary health problems is liver disease. Every year, liver disease claimed the lives of almost 2 million people around the world. There are many different types of liver illnesses, including cirrhosis, hepatitis, liver cancer, and liver tumors. The effectiveness of the strategies was evaluated from a number of angles, including precision, recall, f-1 score, and accuracy. The main perspective for liver illness is one of these truths. Alcohol, smoking, and infections that affect the liver contribute to liver disease. This project's domain is machine learning. Predicting outcomes is the key component efficiently and cut costs in the medical industry. To predict accuracy, machine learning algorithms are applied. Several machine learning techniques, such as Logistic Regression, K-Nearest Neighbor, Decision Tree, Support Vector Machine, Naive Bayes, and Random Forest, have been used. For the purpose of predicting liver disease, various categorization techniques are used. The higher accuracy is predicted using logistic regression [2].

People's health is deteriorating as a result of their habits of eating fast food, consuming alcohol, and smoking cigarettes, and they are dealing with a variety of liver disorders that could endanger their lives. These are the various liver ailments, which also include viral hepatitis, alcoholic liver disease, fatty liver disease, autoimmune liver disease, and genetic liver disease. The previously listed typical liver conditions can result in cancer or liver damage. The predominant disorder is cirrhosis of the liver. Two forms of liver cirrhosis are distinguishable. There are two types of cirrhosis: compensated and decompensated. Machine learning is the project's primary focus. Many machine learning algorithms are used in this research. Predicting the accuracy is our key goal in this situation. It includes machine learning techniques including the Nave Bayes Tree, Support Vector Machine, Decision Tree, and Nave Bayes. Regression and classification problems can both be solved with support vector machines. Support vector machine with improved accuracy is concluded in this paper [3].

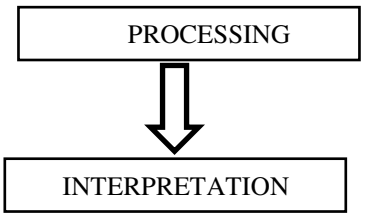


Fig1.1. Process for Liver Disease Prediction

Literature Survey:

Prediction of liver illness using various decision-tree techniques. The proposed technique for categorizing liver illness by this author uses decision trees and includes the following algorithms: J48 (65.69), LMT (69.47), Random Forest (69.30), Decision Stump (70.67), REP tree (66.13), and Random Tree (66.55). The decision stump algorithm's prediction of 70.67 had the highest accuracy out of all of them [1]. a comparison of supervised machine learning algorithms for the prediction of liver disease. The techniques presented by the author are: Naive Bayes 53, K-Nearest Neighbor 62, Support Vector Machine 64, Random Forest 74, Decision Tree 69, Logistic Regression 75. The best predictor of all is Logistic Regression [2].A study on utilizing machine learning to predict liver disease. The k-Nearest Neighbor method (62), Naive Bayes (53), Random Forest (74), Support Vector Machine (64), Decision Tree (69), and Logistic Regression (75) are the

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

algorithms the author proposed in this. Random Forest predicts with the highest degree of accuracy [3]. Using KNN and the Naive Bayes algorithm, the diagnosis of liver illness in patients is predicted. The author's suggested algorithms in this article are K-Nearest Neighbor (72.5) and Naive Bayes (55.4). K-Nearest Neighbor Algorithm has the highest accuracy of all prediction methods [4].

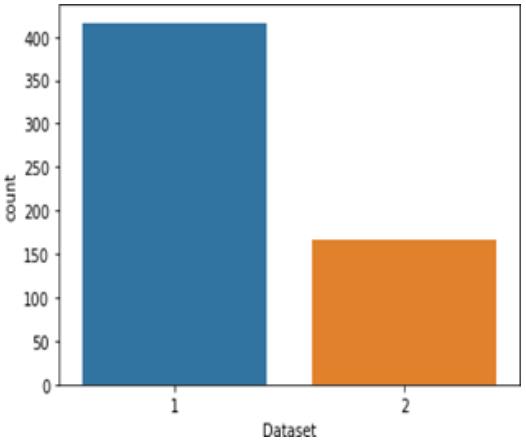
Review of liver disease prediction using machine learning method. Support Vector Machine (74.09), Decision Tree (72.54), Random Forest (72.02), Logistic Regression (71.50), and Gaussian Nave Bayes (57.51) are the algorithms the author proposes in this. Support Vector Machine has the highest accuracy of all prediction methods [5]. system for predicting liver illness utilizing machine learning techniques. Support Vector Machine, Artificial Neural Network, K-Nearest Neighbor Algorithm, and Naive Bayes are the algorithms that the author proposes in this. The highest accuracy of all Support Vector Machine predictions is [6]. use machine learning to classify liver illness. In this, Decision Tree 66, Naive Bayes 56, and Support Vector Machine with Weka Tool 69 are the algorithms that the author proposes. Support vector machine predicts with the highest level of accuracy out of all [7]. Efficiency evaluation of machine learning algorithms for the detection of liver disease. The algorithms Nave Bayes, Decision Tree, Random Forest, and Support Vector Machine are suggested by the author. The highest accuracy of all Support Vector Machine predictions is [8]. Prediction of liver disease using statistical machine learning techniques. Artificial Neural Network (ANN), Support Vector Machine, and Random are the algorithms that the author proposes in this Woods [9]. Prediction of liver disease using machine learning techniques. Support Vector Machine, Artificial Neural Network 70, K-Nearest Neighbor, and Naive Bayes 55.5 are the algorithms that the author proposes in this. The higher accuracy is predicted by Support Vector Machine [10].

Data Set Description

Data Gathering:

A sample data is collected from different internet sources such as Kaggle. In Kaggle we download Indian liver disease dataset which consists of attributes like:

1. Age
2. Gender
3. Total Bilirubin
4. Direct Bilirubin
5. Alkaline phosphatase
6. Alanine Aminotransferase
7. Aspartate Aminotransferase
8. Total proteins
9. Albumin and Globulin Ratio
10. Result.



*Female patients: 142

*Male patients: 441

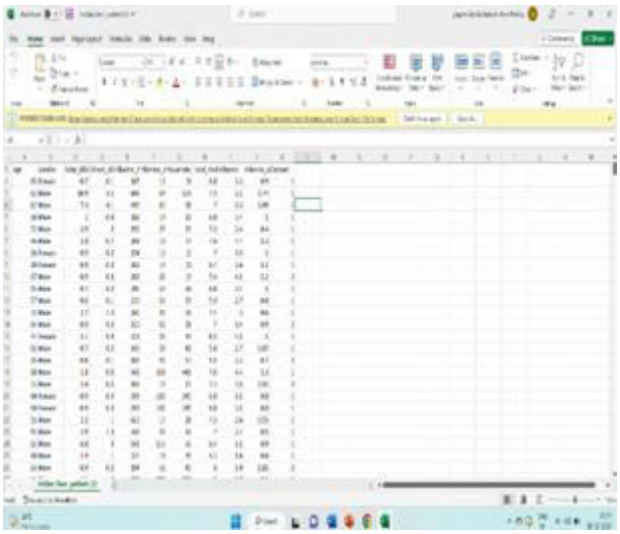
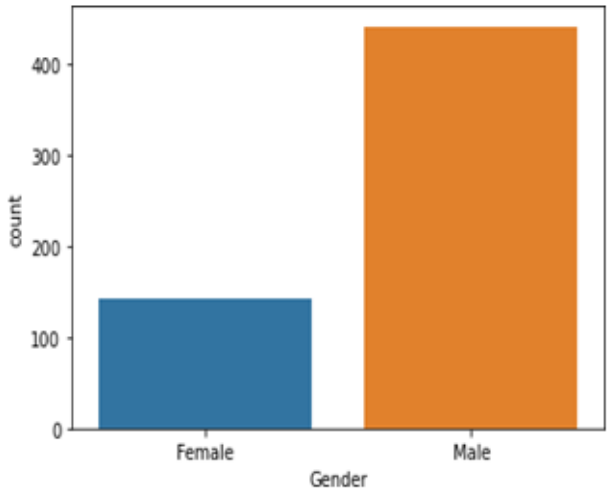


Fig1.2 Indian liver disease Dataset

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Graph Representation



*Number of patients not predicted disease: 416
*Number of patients predicted with disease:167

The dataset consists of different attributes. we are downloading a dataset of historical data who are suffering from liver disease. Finally, the result will be 1 or 2.1 means the person is affected from liver disease.2 means the person is notaffected for liver disease.

Preprocessing:

It is a method for transforming unclean data into clean datasets. The data is collected from several sources in a raw format that makes analysis impractical. Examples of preprocessing include cleaning, transformation, normalization, feature extraction, and selection.

Proposed Methodology:
Supervised Algorithm:
A labelled data set is used to train supervised learning. Regression and classification are both applications of supervised learning. In order to solve challenging problems, supervised learning is frequently used. Learning under supervision yields more accurate results. The purpose of supervised learning is to develop a model that can forecast actual results. The algorithms for supervised learning that are employed are Random Forest and Logistic Regression. Regression and classification both use it.

Random Forest Algorithm
One of the supervised learning algorithms is Random Forest. One of the widely used machine learning algorithms is this one. Both classification and regression can be done with it. It can address machine learning classification issues. To solve a challenging problem, numerous classifiers are combined. This method is used to enhance the model's performance and is an example of ensemble learning. The more trees there are in a forest, the higher the accuracy and the fewer complications there will be. When compared to other algorithms, Random Forest requires less training time. The largest datasets are processed well, and Random Forest predicts results with the highest accuracy. Random forest is a classifier that contains many of decision trees on various subsets of the given dataset and to improve the prediction of accuracy of that dataset.

Logistic Regression:
One of the supervised learning techniques is logistic regression, which may be used for both classification and regression. It introduces discrete parallel items with a range of 0 and 1 and can deal with absolute as well as numerical components. It is employed to forecast the likelihood of the target variable. The target variable has a binary nature (only two sub classes). Simply said, the binary nature of the target variable means that data is encoded as either 1 or 0. A 1 indicates success or a yes, a 0 indicates failure or a no. A model predicts p(Y=1) as a function of X mathematically. It is used to find solutions to issues like diabetes prediction, spam detection, and cancer diagnosis.

The two and three forms of logistic regression are binary or binomial:
There are just two conceivable possibilities in binary or binomial. Either one or zero. These variables could stand for "yes" or "no," "winning" or "losing," "success" or "failure
Multinomial:
Three or more potential unordered kinds or types without any quantitative significance are considered multinomial.
These variables, for instance, might stand in for "type A" or "type B" or "type C".

Ordinal:
Ordinal is defined as three or more unordered types or thetypes having a quantitative significance.
For example: these variables represent as

"poor", "good", "excellent" and each category can have thescores like 0,1,2,3.

Results

The results for all the algorithms and of final outputs areshown in the table:

S.no	Model	Training Score	Test Score	Accuracy
1.	Random Forest	100	78.63	0.7863
2.	Logistic Regression	71.46	76.07	0.7606

Conclusion

Liver-related illnesses are on the rise in frequency. The best approach is to stop smoking, eating junk food, and drinking alcohol because these all worsen liver damage. The major goal of the research is to use machine learning models to predict risk and outcome, such as logistic regression and random forest. We make more accurate predictions using Random Forest.

References

[1] Liver Disease Prediction by using different decision tree techniques international journal of datamining & Knowledge Management process (IJDKP) vol.8, N0.2, March 2018 by Nazmun Nahur and Ferdous Ara.

[2] A Comparative study on Liver disease prediction using supervised machine learning algorithms International Journal of Scientific and Technology research volume 8, issue 11, November 2019.

[3] A study of liver disease prediction using machine learning approaches by MD. Abdur Rahmun Bhuiyan Daffodil International University December 2019.

[4] Prediction for diagnosis of liver disease in patients using KNN and Naive Bayes Algorithm 2nd International conference on cybernetics and intelligent system (ICORIS) 2020.

[5] Review of liver disease prediction using machine learning algorithm Journal of Emerging Technologies and Innovative Research (JETIR) volume 8, February 2021.

[6] Liver disease prediction system using machine learning techniques International Journal of Engineering Research and Technology volume 10, June 2021.

[7] Liver disease prediction using machine learning classification volume 18, special issue on information retrieval and web search September 2021

[8] Efficiency measure of machine learning algorithms on liver disease diagnosis psychology and education, ISSN:00333077 ,2021.

[9] Statistical ML approaches to liver disease prediction Department to mathematics and statistics Texas Tech University December 2021.

[10] Liver Disease Prediction using Machine LearningTechniques volume 10 June 2021.

[11] Liver disease prediction using machine learning algorithms international general of innovative research in technology volume 9, ISSN:2349-6002, June 2022

[12] Prediction of liver disease using machine learning International Journal of Advanced Research in Science Communication and Technology (IJARSCT), volume 2, issue 1, 2022.

[13] statistical machine learning approaches to liver disease prediction International Journal of Scientific Research and Development volume 5, August 2022.

[14] Liver Disease Prediction using Machine Learning Techniques AIP Conference 2022.

[15] Prognosis of liver disease using machine learning algorithms International Conference on Recent Innovation in Electrical ,Electronic and Communication Engineering (ICRIEECE) 2022.