International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.031 ISSN: 2581-4621

Early Detection of Liver Disease by using Machine Learning

Mr. Shibam Ch. Karmakar[1], Mr. Subham Pratihar[2], Ms. Shreeja Roy[3], Mrs. Sulekha Das[4], Dr. Avijit Chaudhuri[5]
[1,2,3]UG Student, 2nd Year, Techno Engineering College, Banipur
[4,5]Assistant Professor, Department of Computer Science and Engineering, Techno Engineering College, Banipur
Author's Orcid ID : [1]0009-0008-8806-1743, [2]0009-0003-6134-0498 [3]0009-0005-2326-6412, [4]0000-0002-6641-3268, [5]0000-0002-5310-3180

Abstract
The liver is the largest internal organ of the human body. It is responsible for conversion of food intake into useful nutrients and also helps to store them. It is responsible for conversion of toxic molecules into harmless particles. But recent studies report significant deaths due to liver diseases. It is mainly due to unhealthy diet habits and unhealthy routine of people. In the race of doing work people are ignoring their health resulting in abnormal health and affecting the liver significantly. Therefore prediction of liver disease with high accuracy and speed is an important concern. The liver tissues undergo deformation or abnormalities comparatively slower than other body tissues, so detection becomes more difficult. In recent decades, the use of automatic decision making systems and tools has found a significant role in the medical field. As the medical field deals with human life, by using the knowledge of machine learning, deep learning, artificial intelligence, and big data we can help in rapid and appropriate treatment and cure. This will help physicians in making the correct decision at the right moment and appropriate procedure. In this regard, this study provides an extensive review of the progress of applying Artificial Intelligence in forecasting and detection liver diseases and then summarizes related limitations of the studies followed by future research.
Keywords: Liver Diseases, Machine learning, Data Mining, Deep learning, Artificial Intelligence.

1. Introduction

Machine learning (ML) techniques help us to make better decisions and distinguish many diseases with accuracy levels. Medical fields produce and collect large volumes of data that can be processed using machine learning to improve the efficiency of patient care, and to reduce the time of treatment. Machine learning has a vital role in medical science as this field deals with human life and well-being. In this dataset a total of 583 records is present, where 416 records are present for liver-disease patients and 167 persons are non-liver patients. The data are collected from test samples by studying the medical test records of patients from North-East of Andhra Pradesh, India and are available in the UCI repository. Out of the 583 records, 441 are male patients and 142 are female patients.
In this paper, a machine learning method is used to predict liver disease, and to find out the performance of prediction accuracy. In this regard and to achieve this aim, a logistic regression algorithm is first produced to predict liver disease in its early stage. This helps the model to achieve better accuracy in the prediction. In the end, the performance of the proposed algorithm is assessed when it applies to a liver database.

2. Methodology

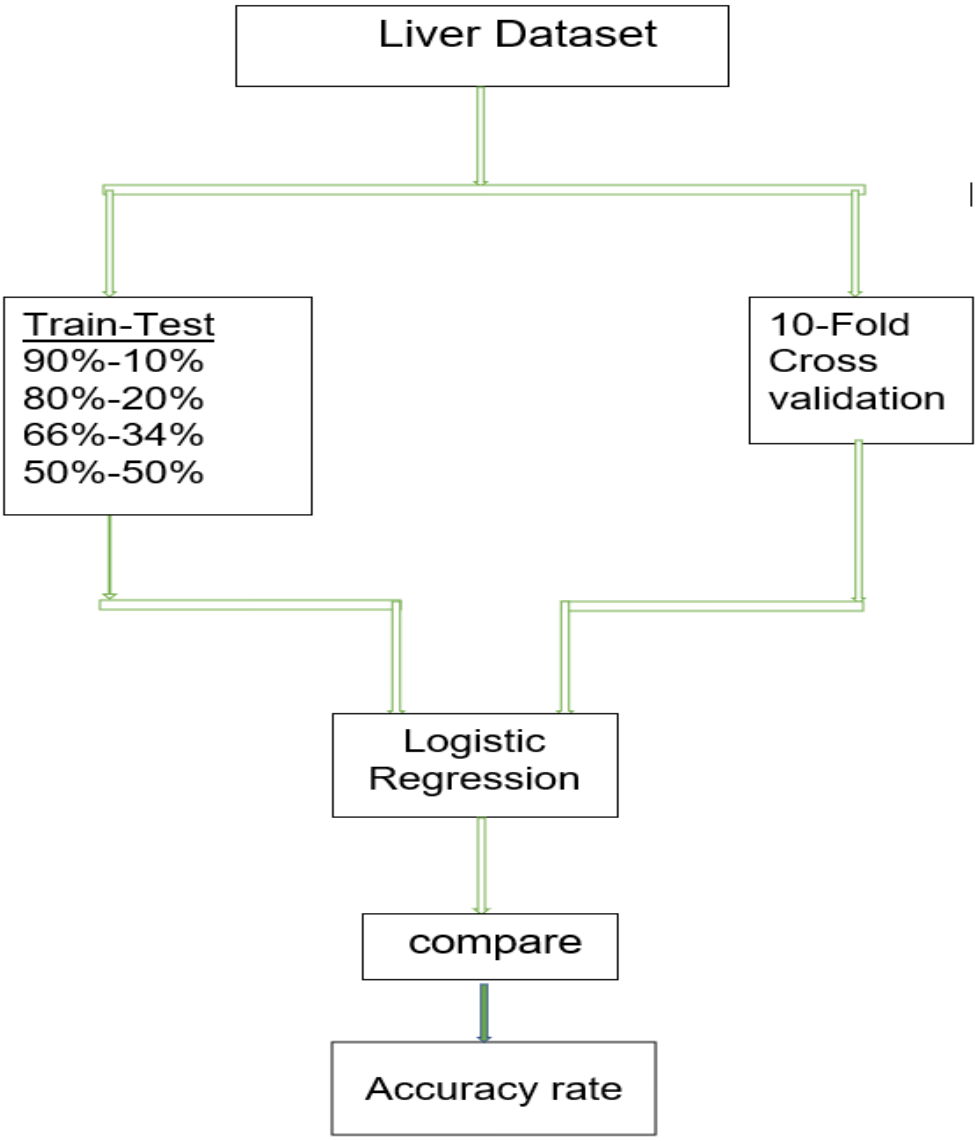
The attributes (independent and dependent variables) on which liver disease depends are listed below:

Attributes	Description
Age	Age of patients
Gender	Gender of patients

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.031 ISSN: 2581-4621

Total Bilirubin	Total bilirubin rate present in patients
Direct Bilirubin	Direct Bilirubin rate present in patients
Alkaline Phosphatase	Alkaline Phosphatase rate present in patients
Alanine Aminotransferase	Alanine Aminotransferase rate present in patients
Aspartate Aminotransferase	Aspartate Aminotransferase rate present in patients
Total Proteins	Total Proteins rate present in patients
Albumin	Albumin rate present in patients
Albumin and Globulin Ratio	Albumin and Globulin Ratio rate present in patients
Dataset	

Flowchart of our work :



International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.031 ISSN: 2581-4621

Multiple Logistic Regression - Multiple Logistic Regression is a machine learning algorithm used to predict a single output which is a binary variable using one or more other variables. It is also used to calculate the numerical relationship between those given sets of variables.

FORMULA OF COEFFICIENT OF MULTIPLE LINEAR REGRESSION:

$$b_i = \frac{(x_1 - \Sigma x_i)(y_1 - \Sigma y_i)}{(x_i - x)^2}$$

The model of multiple regression can be represented as :

$Y = a + b_1X_1 + b_2X_2 + \ldots\ldots\ldots\ldots\ldots + b_nX_n$

Where Y = Dependent Variable (Dataset)

a = Constant Variable

b_1 = Coefficient of first independent variable

b_2 = Coefficient of second independent variable

b_3 = Coefficient of third independent variable

b_4 = Coefficient of fourth independent variable

b_5 = Coefficient of fifth independent variable

b_6 = Coefficient of sixth independent variable

b_7 = Coefficient of seventh independent variable

b_8 = Coefficient of eighth independent variable

b_9 = Coefficient of ninth independent variable

b_{10} = Coefficient of tenth independent variable

X_1 = Independent Variable (Age)

X_2 = Independent Variable (Gender)

X_3 = Independent Variable (Total Bilirubin)

X_4 = Independent Variable (Direct Bilirubin)

X_5 = Independent Variable (Alkaline Phosphatase)

X_6 = Independent Variable (Alamine_Aminotransferase)

X_7 = Independent Variable (Aspartate Aminotransferase)

X_8 = Independent Variable (Total Proteins)

X_9 = IndependentVariable (Albumin)

X_{10} = Independent Variable (Albumin and Globulin Ratio)

The logistic regression is presented as:

$$Y_1 = \frac{Y}{(1 + e^{-Y})}$$

Here,

Y= Dependent Variable e = Euler's number

Logistic regression - Logistic regression is a machine learning algorithm used to check and calculate the relationship between a dependable variable and one or more independent variables. It is a type of regression where a dependable variable is binary.

ACCURACY: Ratio of the correctly classified subjects to the whole subjects'. Accuracy is a measure of prediction.

PRECISION: Ratio of the correctly positive classified by our program to all positive classified.

SPECIFICITY: Ratio of the number of correctly negative classified subjects to the total number of negatives subjects'

SENSITIVITY: Ratio of the number of true positives to the total no. of positives.

- ACCURACY = (TP + TN / TP +TN + FP + FN) * 100
- PRECISION = (TP / FP + TP) * 100

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.031 ISSN: 2581-4621

- SPECIFICITY = (TN / TN + FP) * 100
- SENSITIVITY = (TP / TP + FN) * 100

$$Kappa\ Test\ = \frac{Observed\ Conditions\ -\ Expected\ Conditions}{100\ -\ Expected\ Conditions}$$

where,
Observed Conditions = % of (Overall Accuracy)

$$Expected\ Conditions\ = \frac{(TP\ +\ FP)\ *\ (TP\ +\ FN)\ +\ (FN\ +\ TN)\ *\ (FP\ +\ TN)}{100}$$

3. Results

Table.3.1. Results of 10 fold Cross Validation :

0-58 data as Test Data Confusion Matrix: 41 0 0 17 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0	59-117 data as Test Data Confusion Matrix: 40 0 0 18 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0
118-176 data as Test Data Confusion Matrix: 49 0 0 9 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0	177-235 data as Test Data Confusion Matrix: 42 0 0 16 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0
236-294 data as Test Data Confusion Matrix: 43 0 0 15 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0	295-353 data as Test Data Confusion Matrix: 37 0 0 21 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0
354-412 data as Test Data Confusion Matrix: 38 0 0 20 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0	413-471 data as Test Data Confusion Matrix: 40 0 0 18 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0
472-530 data as Test Data Confusion Matrix: 40 0 0 18 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0	531-589 data as Test Data Confusion Matrix: 44 0 0 14 Accuracy: 100.0 Precision: 100.0 Recall: 100.0 Specificity: 100.0

Table.3.2. Accuracy of difference between Actual Data and Calculated Data :

DOI:10.46647/ijetms.2023.v07i02.031 ISSN: 2581-4621

Accuracy of 90%Data as Training Data or (0.90)	100
Accuracy of 80%Data as Training Data or (0.80)	100
Accuracy of 66%Data as Training Data or (0.66)	100
Accuracy of 50%Data as Training Data or (0.50)	100

4. Conclusion

In this paper a model is proposed where it uses multiple logistic regression for liver disease detection. Secondary data is collected and used from the UCI repository to calculate relationships between dependent and independent variables. We proceed to find a confusion matrix to compare accuracy between actual data and calculated data produced by our model. We then applied 10 - fold cross validation to calculate accuracy, precision, specificity, sensitivity and kappa. We calculated the confusion matrix for each sub - list. This paper will try to produce a new and improved expert system for early detection of liver disease.

5. References

1. Negar Maleki, Yasser Zeinali, Seyed Taghi Akhavan Niaki,A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection,Expert Systems with Applications,Volume 164,2021,113981,ISSN 0957-4174,
2. Shaheamlung, G., Kaur, H., & Kaur, M. (2020, June). A Survey on machine learning techniques for the diagnosis of liver disease. In 2020 International Conference on Intelligent Engineering and Management (ICIEM) (pp. 337-341). IEEE.
3. A. K. Chaudhuri and A. Das, "Variable Selection in Genetic Algorithm Model with Logistic Regression for Prediction of Progression to Diseases," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-6, doi: 10.1109/INOCON50539.2020.9298372.
4. Ramana, Bendi Venkata, M. Surendra Prasad Babu, and N. B. Venkateswarlu. "A critical study of selected classification algorithms for liver disease diagnosis." International Journal of Database Management Systems 3.2 (2011): 101-114.
5. Ramana, Bendi Venkata, MS Prasad Babu, and N. B. Venkateswarlu. "Liver classification using modified rotation forest." International Journal of Engineering Research and Development 6.1 (2012): 17-24
6. Yadav, S., & Shukla, S. (2016, February). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In 2016 IEEE 6th International conference on advanced computing (IACC) (pp.78-83). IEEE.
7. Behera, M. P., Sarangi, A., Mishra, D., & Sarangi, S. K. (2023). A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine. Procedia Computer Science, 218, 818-827.
8. Mehtaj Banu H" Liver Disease Prediction using Machine-Learning Algorithms" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6, August 2019
9. Durai, Vasan, Suyan Ramesh, and Dinesh Kalthireddy. "Liver disease prediction using machine learning." (2019).
10. https://www.worldlifeexpectancy.com/life-expectancy-research
11. Sivakumar D , Manjunath Varchagall , and Ambika L Gusha S "Chronic Liver Disease Prediction Analysis Based on the Impact of Life Quality Attributes." (2019). International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume7, Issue-6S5, April 2019
12. https://scholar.google.com/

DOI:10.46647/ijetms.2023.v07i02.031 ISSN: 2581-4621

13. https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence#:~:text=Artificial%20intelligence%20is%20the%20simulation,speech%20recognition%20and%20machine%20vision.
14. https://www.sciencedirect.com/science/article/pii/S095741741630464X
15. https://www.researchgate.net/profile/Animesh-Samanta-3/publication/360033886_Prediction_through_machine_learning_on_the_dependence_of_job_prospects_in_the_Afro-American_community_on_proficiency_in_English/links/625e86e79be52845a90f132c/Prediction-through-machine-learning-on-the-dependence-of-job-prospects-in-the-Afro-American-community-on-proficiency-in-English.pdf