

# Statistical Machine Learning Approaches To Liver Disease Prediction

## LITERATURE SURVEY:

### 1) Prognosis of Liver Disease: Using Machine Learning Algorithms

Data mining classification techniques like Decision Tree, Linear Discriminant, SVM Fine Gaussian and Logistic Regression algorithms are applied. Laboratory parameters of the patients are used as the dataset. Data contains features that can establish a rigorous model using Classification technique. MATLAB2016 is used in this paper for implementing classification algorithm on the dataset. Linear Discriminant algorithm showed the highest prediction accuracy 95.8% and ROC is 0.93. This paper helps in foresee the presence of liver disease using different classification algorithms. The dataset considered consists of lab reports of 574 patients who are advised with LFT. The initial phase includes assorting the information from the available source. This information is of the patients who are predicted of having liver disorder. The dataset is analyzed in the next phase to check which algorithm can be applied to obtain optimal results. The last phase includes classification of data by application of proper algorithm on the dataset and formatting a training model. In the last phase data is tested against the trained model to get proper predictive values. The proposed paper makes use of laboratory parameters of individuals after taking LFT. The model that yields the maximum efficiency is considered. Once the patient is predicted with the liver disease, he is further advised for imaging tests to determine the existence of tumor lesions and their stage. This paper makes use of the lab test reports of the patients who has undergone Liver Function Test. MATLAB2016 is used but Logistic Regression gave high accuracy of 95.8%. Various predictors are tested by plotting graph that determined the existence of disorder in liver. Further research is proposed for considering the tumor characteristics of the patient once he is diagnosed with liver disorder. Also large dataset can be considered for training the model and algorithms can be determined.

### 2. Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms

C.Geetha (Assistant Professor) Department of CSE, Bharath Institute of Higher Education and Research Chennai, India

Geetha, Department of CSE Dr. MGR University

Centered on their success variables, this research also aims to compare the classification algorithms and to provide prediction accuracy results.

In order to store large-scale information on patient outcomes, procedures, etc. Electronic health records (EHR) are used. The data on the EHR can be organised or unstructured. Electronic health records are stored in a standardised data format using managed language to log patient knowledge as a written text that is hyperlinked in existence. The EHR aims to streamline knowledge about the clinical workflow. Ensemble learning is a well-known method used for prediction by integrating multiple ensemble models of machine learning. Aggregations of various classifiers are J48, C4.5 and Naive Bayes. Ensembles search for better outcomes than all of the simple classifiers. The proposed work aims to enhance the predictive and classification quality of healthcare data by developing a hybrid predictive classifier model using the classifier ensemble. The SVM, Logistic Regression, comprises two main machine learning techniques. Using all the models, the prediction analysis has been implemented and their performance has been assessed. The probability of liver disease prediction attained with an accuracy of 96%. In future, the present scenario can be compared with other techniques such as naive bayes's classification, Random forest etc. Also this work can be further focused on implementation of parametric classifications by bio-inspired optimization algorithms.

### **3.Liver Disease Prediction System using Machine Learning Techniques**

1.Rakshith.D B, 2.Mrigank Srivastava

3.Ashwani Kumar

4.Gururaj.S P,From the Department of Computer Science and Engineering, Siddaganga Institute of Technology,Tumkur,India.

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. In this project we have taken UCI ILPD Dataset which contains 10 variables that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos and contains 415 as liver disease patients and 167 as non-liver disease patients. As we go through the next parts of this paper we will explain what process has taken place for the selection of best model and building necessary system for the prediction of liver disease. The major outcomes that can be expected through this project is that it increased convenience for predicting a liver disease. So, in such a scenario, our project will be extremely helpful to the society. With the dataset that we used for this project, we got 100 % accuracy for SVM model, and though it might be difficult to get such accuracies with very large datasets, from this project's results, one can clearly conclude that we can predict the risk of liver diseases with accuracy of 90 % or more.

### **4.LIVER DISEASE DIAGNOSIS USING MACHINE LEARNING**

1.Prof. Sayalee Deshmukh , 2.Anushka Sawant,

3.Manasi Khopade,4.Pratiksha Kawale,

5.Yashika Palan,

Department of Computer Engineering, Bharati Vidyapeeth's College of Engineering for Women, Pune-411043

The major goal of this study is to employ classification algorithms to distinguish between liver patients and healthy people. Chemical components (bilirubin, albumin, proteins, alkaline phosphatase) present in the human body, as well as tests such as SGOT and SGPT, determine whether a person is a patient, or whether they need to be diagnosed. Excessive alcohol consumption, inhalation of toxic gases, eating of contaminated food, pickles, and medicines have all contributed to an increase in patients with liver disease. The goal of this research is to analyse prediction algorithms in order to relieve doctors of their workload. Keywords— liver disease, SVM, Random Forest, KNN, ML, python, etc. The major goal of this study is to employ classification algorithms to distinguish between liver patients and healthy people. In this study, the performance of FIVE classification techniques was compared using data from liver patients: Logistic Regression, Support Vector Machines (SVM), K Nearest Neighbour (KNN), Decision Tree and Random Forest (RF). Furthermore, the most accurate model is implemented as a user-friendly Graphical User Interface (GUI) in Python using tkinter package. Doctors and medical practitioners can easily use the GUI as a screening tool for liver disease. The dataset used in this work is The Indian Liver Patient Dataset (ILPD), which was chosen from the UCI Machine Learning repository. It is a representative sampling of the entire Indian population.

## **5) A Survey on machine learning techniques for the diagnosis of liver disease**

1. Golmei Shaheamlung ,
2. Harshpreet , 3. Mandeep.

The motive of this paper is to give a survey and comparative analysis of the entire machine learning techniques for diagnosis and prediction of liver disease in the medical area, which has already been used for the prediction of liver disease by various authors and the analysis are based on Accuracy, Sensitivity, Precision, and Specificity. With the help of this survey and study, it has clearly found and observed that some machine learning algorithm such as Decision tree, J48 and ANN provide better accuracy on detection and prediction of liver diseases. With this in mind and performance we know that different algorithm has different performance based on different scenario but most importantly, the dataset and feature selection is also very important to get better prediction results. With this survey we found out that the accuracy and performance can be improved by using different combination or hybrid machine learning algorithm and in future we can also work on more parameter's which help to get better performance than the existing technique.

## **6) Prediction of Liver Disease using Classification Algorithms**

1. Thirunavukkarasu.K,
2. Ajay S. Singh, 3. Md Irfan,
4. Abhishek Chowdhury

The main aim is to predict liver disease using different classification algorithms. The algorithms used for this purpose of work is Logistic Regression, K-Nearest Neighbour and Support Vector Machines. Accuracy score and confusion matrix is used to compare this classification algorithm. In this paper, Logistic Regression, K-Nearest neighbour and Support Vector Machines are been used for prediction of liver disease. The proposed methods are used to compare classification accuracy of Logistic Regression, K-nearest neighbour and Support Vector Machine. The comparison of all these algorithms been done are based on classification accuracy which is found through confusion matrix. From the experiment, Logistic Regression and K-Nearest Neighbour have the highest accuracy but logistic regression have the highest sensitivity. Therefore, it can be concluded that Logistic Regression is appropriate for predicting liver disease.

## **7) Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques**

1. Sateesh Ambesange, 2. Vijayalaxmi A,
3. Rashmi Uppin, 4. Shruthi Patil, 5. Vilaskumar Patil

Random Forest (RF) algorithm is used to predict the disease with different preprocessing techniques. Data set is checked for skewness, outliers and imbalance using univariate and bivariate analysis and then suitable algorithms used to remove outliers and various oversampling and under sampling techniques are used to balance such data's. In this work, Indian Liver Patient Dataset hosted at "ics.uci.edu" is used. 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 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. In this work, ML models are built using various preprocessing techniques to balance the unbalanced data and predicted using RF algorithm.