International Journal of Research in Engineering, Science and Management
Volume 7, Issue 4, April 2024
https://www.ijresm.com | ISSN (Online): 2581-5792

Predictive Modelling for Liver Disease Diagnosis

Sameena Bano[1*], Salma Jabeen[2], Mohammed Kaleem[3]
[1]*Associate Professor, Department of Computer Science and Engineering, Don Bosco Institute of Technology, Bangalore, India*
[2]*Associate Professor, Department of Information Science and Engineering, Don Bosco Institute of Technology, Bangalore, India*
[3]*Assistant Professor, Department of Computer Science and Engineering, Don Bosco Institute of Technology, Bangalore, India*

***Abstract*: In this project, patient data sets are investigated to see if it is possible to predict whether a subject will acquire liver disease using simply a rigorous classification model. Since there are already procedures in place to analyze patient data and classifier data, the most crucial aspect in this situation is to anticipate the same decisive outcome with a higher level of accuracy. Recent investigations on the diagnosis of the liver revealed differences in the classification accuracy of different classifiers using varied data sets. The K-Nearest Neighbor Classifier is seen to be providing the best outcomes with India's complete feature set of liver patient data combinations. For the India liver dataset, performance is improved in comparison to the complete UCLA liver dataset and particular algorithms. In order to understand the cause. We recommend examining the liver to account for this disparity. Patients from both India and the USA To date, thorough ANOVA and MANOVA analyses have been performed on these data sets to spot any notable variations between the groupings. It has been noted that people with liver problems. The two nations have a lot of differences.**

***Keywords*: K-Nearest Neighbor Classifier, UCLA, Patients.**

1. Introduction

The liver is probably the vastest organ in the entire body of a person. and it is in charge of processing food and releasing toxic toxins. Viruses and alcohol consumption both induce liver damage, which can endanger a person's life. The liver is affected by hepatitis, cirrhosis, liver tumors, liver cancer, and a range of other disorders. It is in charge of various vital life activities, including bile synthesis for digestion, blood cleansing, blood toxicity management, bilirubin clearance, body metabolism, and the conversion of dangerous ammonia to urea. Liver disease is defined as any abnormality in the liver's function that causes illness. Liver disease, which is caused by the accumulation of fat in the liver, is a highly frequent ailment in India, with over 10 million cases recorded each year. As a result of absence of damaging implications, testing is essential for diagnosis. Artificial Intelligence (AI) is a sort of computerized reasoning that is enabled by computer programmers' ability to learn, gain knowledge, and then use that knowledge in a variety of domains. Man-made awareness is now employed in practically every field, and it is made up of numerous components such as deep learning and machine learning. Obesity, inhalation of poisonous gases, ingestion of polluted food, excessive use of foods and pharmaceuticals, and alcohol are all key causes of liver disease. The intent of the present investigation is to supply machine learning algorithms based on the Classification of Liver Disorders to alleviate doctors efforts.

2. Literature Survey

A. A Critical Comparative Study of Liver Patients: An Exploratory Analysis

In this research, standard statistical approaches such as one-way ANOVA and multivariate ANOVA are used to assess the importance of two populations for improved classification. ANOVA is used to test the significant difference in a single dependent variable spanning a minimum of two groups formed by single independent or classification variable, whereas MANOVA is used to test the significant difference. This work used two liver patient datasets, one from the Indian state of Andhra Pradesh and the other from the University of California at Irvine (UCI) Machine Learning Repository. Age, gender, total bilirubin, direct bilirubin, alkphos, SGPT, SGOT, total proteins, albumin, and A/G ratio were the characteristics of the Indian data set. Mcv, Alkphos, SGPT, SGOT, and Gammagt were properties of the UCI data set. Alkphos, SGPT, and SGOT were the common liver functional tests in both data sets [1].

B. Liver Disease Prediction Using Machine Learning

The research on the prediction and prevention of Liver Disease using Data Mining and artificial intelligence principles is particularly relevant in this decade. Artificial inelegancy notions are crucial in this regard. Many researchers must use machine learning (ML) models to predict diseases. In this study, we propose empirical statistical analysis to prevent LD and use effective ML models for early prediction of liver disorders at a low cost. During the period 2018-2020, data was collected from hospitals and reputable clinical centers in Andhra Pradesh, India. The data set includes both personal and clinical information. For predicting LDs, we use five well-known machine learning models: KNN, SVM, RF, Nave Bayes, and AdaBoost. For predicting LDs, experimental models with an accuracy of 1 (100%) were used. The remaining performs well as well, with accuracy scores above 0.86 (86%) [2].

*Corresponding author: drsameenakhayum@gmail.com

C. Utilizing Machine Learning for Liver Disease Prediction: A Comprehensive Approach

Every year, liver problems kill millions of people. Viral hepatitis alone kills 1.34 million people. Problems with the liver are difficult to detect early on since it will continue to operate normally even if it is partially destroyed. An early detection of liver disorders improves patient survival. Indians are at a significant risk of having a liver failure. By 2025, India is anticipated to be the World Capital for Liver Diseases. The prevalence of liver infection in India is attributed to sedentary lifestyles, increasing alcohol intake, and smoking. With ill youths, we cannot expect a developed and affluent nation.

In this experiment, we used the UCI ILPD Dataset, which includes 10 variables such as age, gender, total bilirubin, direct bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT, and Alkphos, and 415 individuals with liver disease and 167 without. As we progress through this article, we will discuss the process of selecting the optimal model and developing the necessary system for the prediction of liver illness [3].

D. Liver Disease Prediction by Using Different Decision Tree Techniques

The liver is involved in numerous biological activities, including protein synthesis and blood clotting, as well as cholesterol, glucose (sugar), and iron metabolism. It has a variety of tasks, including the removal of pollutants from the body, and is essential for survival. The lack of these functions might cause serious harm to the organism. When a person's liver is infected with a virus, wounded by chemicals, or attacked by their own immune system, the core hazard is the same: the liver will become so damaged that it can no longer function to keep them alive. Liver illness caused by hepatotropic viruses places a significant strain on health-care resources. Chronic liver disease is caused by hepatitis B virus (HBV), hepatitis C virus, and hepatitis delta virus infections. Acute and chronic liver disease are the most fundamental classifications. The duration of acute liver illness is six months or less. The majority of instances of acute liver illness are caused by acute viral hepatitis or medication responses. The intent of this investigation is to evaluate the decision tree algorithms J48, LMT, Random Tree, Random Forest, REPTree, Decision Stump, and Hoeffding Tree in diagnosing liver illness. The liver dataset is analyzed using the aforementioned decision tree methods, and their performance is compared to seven performance indicators (ACC%, MAE, PRE, REC, FME, and Kappa Statistics and runtime) [4].

E. Liver Disease Prediction using SVM and Naïve Bayes Algorithm

The Support Vector Machine (SVM) technique is a basic yet effective Supervised Machine Learning approach that may be used to generate both regression and classification models. The SVM method works effectively with both linearly and non-linearly separable datasets. Even with a small quantity of data, the support vector machine method performs admirably. The goal of the SVM method is to identify a hyperplane in an N dimensional space that clearly classifies the input points. The extent of the hyper plane depends on by the quantity of facets. If There are two specifications for input, and hyperplane is simply a line. If the number of input characteristics is three, the hyperplane transforms into a two-dimensional plane. In brief, a hyperplane is a (n-1)-D plane with n characteristics [5].

3. Methodology

Clinical evaluations, laboratory testing, and in certain cases, machine learning approaches are used to predict liver disease. Here's a broad approach of predicting liver disease.

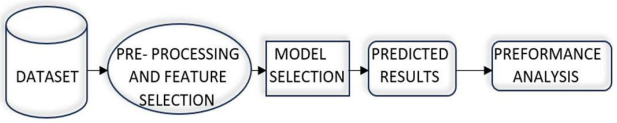

Fig. 1. Architecture diagram

These are the modules that are utilized in prediction.

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyze and Prediction

1) Data Collection

Collecting data is the first genuine step towards the true creation of a machine learning model. This is a vital stage that will determine how successful the model is; the more and better data we collect, the better our model will perform. There are numerous methods for collecting data, such as online scraping, manual interventions, and so on. The entire Ensemble Technique serves to anticipate liver damage.

2) Dataset

The dataset consists of 583 individual data. This data set has 11 components.

- Age: Age of the patient
- Gender: Gender of the patient
- TB: Total Bilirubin
- DB: Direct Bilirubin
- Alkphos: Alkaline Phosphatase
- Sgpt: Alamina Aminotransferase
- Sgot: Aspartate Aminotransferase
- TP: Total Proteins
- ALB: Albumin
- A/R: Albumin and Globulin Ratio
- class:1 Liver diseases & 0 no diseases

3) Data Preparation

Gather data and prepare it for training. Clean up everything that needs it (remove duplicates, rectify mistakes, deal with missing numbers, normalization, data type conversions, and so on). Randomize data, which removes the influence of the order in which we acquired and/or otherwise prepared our data. Visualize data to assist in detecting meaningful associations between variables or class imbalances (bias alert!), or undertake other exploratory research. Divided into training and assessment sets.

Table 1
Dataset overview

	Age	Gender	TB	DB	Alkphos	Sgpt	Sgot	TP	ALB	A/R	Class
1	19	female	0.7	0.1	0.90	6.8	3.3	187	16	18	1
2	76	female	10.9	5.5	0.74	7.5	3.2	699	64	100	1
3	17	male	7.3	4.1	0.89	7.0	3.3	490	60	68	1
4	65	male	1.0	0.4	1.0	6.3	3.4	182	17	20	1
5	68	female	3.9	0.1	0.40	7.3	2.4	195	27	59	1

4) Model Selection

We employed the Gradient Boosting Classifier + Ada Boost Classifier (Ensemble Technique) machine learning technique and obtained an accuracy of 92.1% on the test set, thus we applied this algorithm.

Ensemble Technique:

Ensemble methods are techniques Designed to raise the level of accuracy of findings in models by multiple models instead of using a single model combining the coupled models considerably improve the veracity of the results. As consequence, ensemble approaches in machine learning are becoming in popularity.

5) Analyze and Prediction

This sorting model is assessed by a range of precautions. Accuracy is the ratio of correctly predicted samples to the total number of samples, and classification error is the amount of error present in the samples.

4. Conclusion

Our objective is to determine whether a person has liver disease symptoms based on available information. We are exempting to forecast the sickness by utilizing the Ensemble Technique. This technology would be extremely beneficial to many hospitals and even professional physicians in detecting sickness. In addition, the general public can utilize this technology to diagnose diseases. This method will transform the way things are done and save people's lives as soon as feasible. This entire project is centered on how we can forecast disease using supplied datasets, which will aid in the prevention and treatment of patients' diseases.

References

[1] Bendi Venkata Ramana, Surendra. Prasad Babu. M, Venkateswarlu. N.B,"A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis", International Journal of Database Management Devices, vol. 3, no. 2, pp. 101-114, May 2011.

[2] Vasan Durai, Suyan Ramesh, Dinesh Kalthireddy, “Liver disease prediction using Machine Learning”, International Journal of Advance Research, Ideas and Innovations in Technology, vol. 5, no. 2, 2019.

[3] Rakshith D. B., Mrigank Srivastava, Ashwani Kumar, Gururaj S. P., “Utilizing Machine Learning for Liver Disease Prediction: A Comprehensive Approach”, International Journal of Engineering Research & Technology, vol. 10, no. 6, June 2021.

[4] Nazmun Nahar and Ferdous Ara, “Liver Disease Prediction by Using Different Decision Tree Techniques,” International Journal of Data Mining & Knowledge Management Process, vol. 8, no. 2, March 2018.

[5] S. Vijayarani, S. Dhayanand, “Liver disease prediction using SVM and Naïve Bayes algorithm,” International Journal of Science, Engineering and Technology Research, vol. 4, no. 4, April 2015.