International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 6, Issue 12, ISSN No. 2455-2143, Pages 247-251
Published Online April 2022 in IJEAST (http://www.ijeast.com)

LIVER DISEASE DIAGNOSIS USING MACHINE LEARNING

Prof. Sayalee Deshmukh[#1], Anushka Sawant[*2], Manasi Khopade[#3], Pratiksha Kawale[#4], Yashika Palan[#5]
[#]Department of Computer Engineering,
Bharati Vidyapeeth's College of Engineering for Women, Pune-411043

Abstract— **It is critical to diagnose liver illness early on in order to receive the best therapy possible. Because of the modest symptoms, medical professionals find it difficult to forecast the disease in its early stages. Symptoms frequently appear when it is too late. To address this problem, our research will use machine learning to improve liver disease diagnosis. 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.**

I. INTRODUCTION

Problems with the liver are difficult to detect early on since it will continue to operate normally even if it is partially destroyed. The chances of a patient surviving a liver disease are better if they are diagnosed early. Indians are at a higher risk of liver failure. India is anticipated to become the World Capital for Liver Diseases by 2025. In India, a deskbound lifestyle, increased alcohol intake, and smoking are all factors contributing to the prevalence of liver infection. There are over a hundred different forms of liver infections. As a result, inventing a machine that would aid in disease identification will be extremely beneficial in the medical industry. These technologies will assist physicians in making accurate patient decisions, and with the use of automatic classification tools for liver illnesses (likely mobile enabled or web enabled), the patient wait at liver experts such as endocrinologists will be reduced.

Classification techniques are widely used in medical diagnosis and disease prediction. According to Michael J So rich [1,] the SVM classifier has the best predictive performance for chemical datasets. For the CDC Chronic Fatigue Syndrome dataset, Lung-Cheng Huang found that the Nave Bayesian classifier outperforms SVM and C 4.5. According to Paul R Harper [2], there is no single optimum classification tool; rather, the best performing method is determined by the dataset's characteristics. 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.

Common Liver Disorder

- Hepatitis usually induced by using a virus unfold by using excess infection or direct contact with infected body fluids.
- Fatty liver is a revocable circumstance where giant vacuoles of triglyceride fats gather in liver cells by using the process of limit. It can occur in people with an excessive stage of alcohol consumption as well as in humans who never had alcohol.
- Liver cancer. The risk of liver most cancers is higher in those who have cirrhosis or who had valid types of viral hepatitis; however more often, the liver is the site of secondary (metastatic) cancers spread from other organs.
- Cirrhosis of the liver is one of the most serious liver diseases. It is an action used to point out all types of diseases of the liver characterized via the significant loss of cells. The liver regularly contracts in dimension and becomes leathery and hard. The regenerative action continues below liver cirrhosis however the revolutionary loss of liver cells exceeds cell phone replacement.

II. LITERATURE SURVEY

Health care and medicinal drug handles huge information on every day basis. This information involves of records about the patients, prognosis reviews and clinical images. It is essential to utilize this fact to decipher a decision assist system. To achieve this, it is important to discover and extract the knowledge domain from the raw data. It is accomplished by knowledge discovery and data mining (KDD) [13]. The implementation of facts mining techniques is vast in biological domain. In recent years, liver problems have excessively improved and liver ailments are turning into one of the most deadly illnesses in a number of countries. In this study, liver patient datasets are look at for constructing classification fashions in order to predict liver disease. Several function model development and comparative analysis are carried out for improving prediction accuracy of Indian liver patients. Different studies have been performed for classification of liver disorders.

Classification algorithm is one of the greatest extensive and relevant statistics mining methods used to apply in sickness prediction. Classification algorithm is the most common in a number of automatic medical health diagnoses. Many of them shows excellent classification accuracy.

In [3] this paper, different machine learning algorithms are used such as the methods of Support Vector Machines (SVM), Decision Tree (DT) and Random Forest (RF) are proposed to predict liver disease with better precision, accuracy and reliability. [12] The important goal of this paper is to predict liver ailments the usage of different classification algorithms. These classification algorithms are compared primarily based on the overall performance.

Nazmun Nahar and Ferdous Ara et al., [2] carried out decision tree algorithms: J48, LMT, Random Forest, REP Tree, Decision Stump and Hoeffding Tree to predict the liver disease. A comparative study find out about additionally has been carried out amongst these algorithms. The system analyses the performance of all the algorithms by using measuring their accuracy, F-measure, precision, recall, suggest absolute error.[4] This paper compares various classification models and visualization techniques used to predict liver disease with feature selection. Study and analysis of liver disease prediction has been done. Genetic algorithm combined with XG Boost which is used to fetch the best attributes required for prediction of liver disease.

In [17] this work to build the machine-learning model, Indian Liver Patient Dataset is used, which is based on Indian patient and Random Forest (RF) algorithm is used to predict the disease with different pre-processing 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 the data. Further refinement of model is done through hyper parameter tuning using grid search and feature selection. The final model provides 100% accuracy and also good score across different metrics.

[5] This paper presents a deep-learning-based framework for the segmentation of vacuoles in liver images and also study the correlation of automated quantification with expert pathologist’s manual evaluation. [16]Mohammad Badri Tamam uses Naïve Bayes and KNN algorithms to solve predictive problems based on the results of testing for patients with liver disease or not using the python application.

This article [18] discusses different data mining algorithms like K-Nearest Neighbour (KNN), Decision Tree (DT) and Adaptive Neuro-Fuzzy Inference System (ANFIS) that are used to provide a decision support model that could help the physician in predicting the liver disease from the dataset. The performance of each algorithm is evaluated with respect to accuracy, sensitivity, precision and specificity. A survey on the efficiency of these algorithms is presented.

In this paper [14], G. Shobana proposed a method of feature reduction using Recursive Feature Elimination and applying the Machine learning boosting algorithms to enhance the prediction accuracy. Basic machine learning models were applied to the dataset, where Logistic regression and Multi-Layer Perceptron had higher prediction accuracies with reduced features. Boosting algorithms like Cat Boost, LGBM Classifier, XG Boost and Gradient Boost were applied to the dataset. The impact of feature reduction was investigated on the Gradient boosting machine learning algorithms.

[6] Some of the supervised machine learning techniques used in this for the prediction of heart disease. [7] propose a real-time heart disease prediction system using ML. [8] This paper presents a majority voting ensemble method that is able to predict the possible presence of heart disease in humans [9] This review intends to provide a comprehensive survey of currently proposed machine learning based dental disease detection systems [10] In this research study, the effects of using clinical features to classify patients with chronic kidney disease by using support vector machines algorithm is investigated. [11] This paper proposes an electromagnetic system, including an antenna operating across the band 0.4-1 GHz as a data acquisition device and a supervised Machine Learning (ML) framework to learn an inferring model for FLD directly from collected data.

[15] Dr. Vijayalakshmi M.N uses the lab test reports of the patients who has undergone Liver Function Test. MATLAB2016 is used and developed a model by applying classification algorithms SVM, Logistic Regression and Decision tree. Logistic Regression gave high accuracy of 95.8%. Various predictors are tested by plotting graph that determined the existence of disorder in liver.

Inference of Literature Survey:
The purpose of this literature surveys is to understand how a problem can be solved with the help of various methods. Lastly, we conclude that the literature survey helped us to know the various methods to diagnose liver disease. Here, in this proposed model we will be using Random Forest (RF) algorithms in order to diagnose the liver disease since it provides us with high accuracy.

III. METHODOLOGY

A. Proposed System

In this proposed framework, we use dataset which consists of Indian liver patient data. Making use of this dataset, we carry out pre-processing and feature selection on this particular dataset. At this point, we have enormous number of features in dataset; hence feature selection is a vital part in our Machine Learning model. As we use feature selection, it offers us with most significant features and this in turn makes our disease diagnosis model more accurate and provides us with better results.

Hence, we will be using Logistic Regression, Support Vector Machines (SVM), K Nearest Neighbour (KNN), Decision Tree and Random Forest (RF) algorithms. We take the help of Random Forest Classifier algorithm to train our model for liver disease diagnosis. Furthermore, the model is implemented as a user-friendly Graphical User Interface (GUI).

B. System Architecture

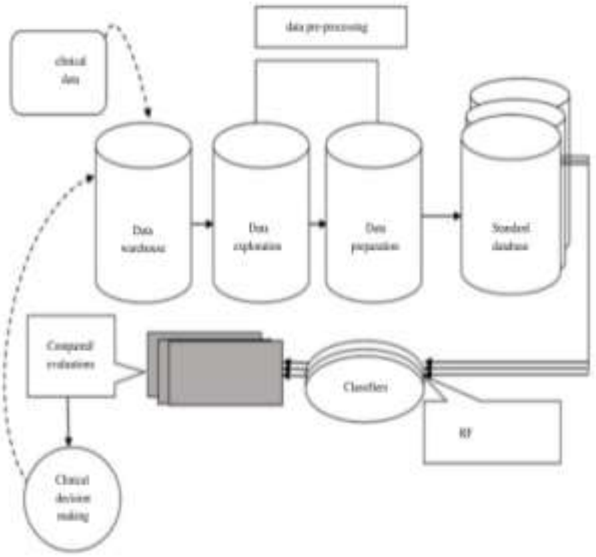

Fig. System Architecture

Explanation:

1) Initially, we will take clinical data (Indian Liver Patient Dataset) as input for liver disease diagnosis model. Indian Liver Patient Dataset (ILPD) was obtained from the UCI Machine Learning Repository. There have been 583 cases based totally on ten distinctive biological parameters in the dataset. Based on these criteria, the class value was once stated as either yes (416 cases) or no (167 cases), reflecting the liver.

2) **Data Exploration**:
Initial step of statistics evaluation which inculcates summarizing the data and observing initial patterns in the data and attributes is known as Data Exploration. Various visualization strategies such as histogram and boxplot are used to discover the extreme and outlier values. Feature correlation of values is assessed in order to identify linearly dependant features.

3) Data Pre-processing:
The data may have missing values and outliers. So, this can be removed by data pre-processing. Data pre-processing is the process of transforming raw data into an understandable format. It is also an necessary step in data mining as we cannot work with raw data. Pre-processing of data is primarily to check the data quality. The major tasks involved in data pre-processing:

- Data cleaning
- Data integration
- Data reduction
- Data transformation

4) Data Preparation:

- ➢ Imputation of missing values: It refers to identifying missing values in the data and imputing the empty values with median values.

- ➢ Elimination of duplicate values: In order to improve the efficiency and quality of data.
- ➢ Outlier detection and Elimination: Outliers are extreme values that significantly deviate from the rest of the values which is caused due to inappropriate measurement or experimental error.

The next step is feature selection. Feature selection is basically the part where in we reduce the number of input variables to those that seem most useful in a way of predicting our target Variable.

5) Classification Algorithms:
Classification algorithm is one of the greatest significant and applicable data mining techniques used to apply in disease prediction. Classification algorithm is the most common in several automatic medical health diagnoses. Many of them show good classification accuracy. Different data mining algorithms like Logistic Regression, Support Vector Machines (SVM), K Nearest Neighbour (KNN),

Decision Tree and Random Forest (RF) were implemented. The algorithms are briefly discussed below:

- **Random Forest:**

Random Forest algorithm is a supervised classification algorithm. We can see it from its name, which is to create a forest by some way and make it random. There is a direct relationship between the number of trees in the forest and the results it can get: the larger the number of trees, the more accurate the result. But one thing to note is that creating the forest is not the same as constructing the decision with information gain or gain index approach. They are classifiers that construct decision trees for training input. A random value is assigned as range to feature space for splitting the tree. Based on the training ensemble class value is predicted as the modal value of distinct tree.

Now, we have trained our model. We can use this model after evaluation to diagnose the liver disease.

IV. CONCLUSIONS

With the passage of time, diseases of the liver and heart are becoming increasingly widespread. These are solely going to get worse in the future, thanks to ongoing technological improvements. Despite the truth that people are becoming more health-conscious and enrolling in yoga and dancing classes, the sedentary lifestyle and facilities that are continuously being delivered and improved will proceed to be an issue. As a result, in this situation, our project will be rather recommended to society.

V. REFERENCES

[1] VasanDurai, Dinesh, Kalthireddy, “Liver disease prediction using machine learning”, International Journal of Advance Research, Ideas and Innovations in Technology, 2017.

[2] Nazmun Nahar and Ferdous Ara, “Liver disease prediction by using different Decision tree techniques”, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.8, No.2, March 2018.

[3] A.Sivasangari,Baddigam Jaya Krishna Reddy, Annamareddy Kiran, P.Ajitha, “Diagnosis of Liver Disease using Machine Learning Models,” IEEE Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020.

[4] Maria Alex Kuzhippallil, Carolyn Joseph, and Kannan A, “Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients,” 6th International Conference on Advanced Computing & Communication Systems (ICACCS), 2020.

[5] Sanket Deshmukh, AvinashLokhande, RatulWasnik, and Nitin Singhal, “Vacuole Segmentation and Quantification in Liver Images of Wistar Rat,” 978-1-7281-1990-8/20/$31.00 IEEE, 2020.

[6] Rahul Katarya, Polipireddy Srinivas, “Predicting Heart Disease at Early Stages using Machine Learning: A Survey”, International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020.

[7] Abderrahmane Ed-daoudy*, Khalil Maalmi, “Real-time machine learning for early detectionof heart disease using big data approach”, 2019.

[8] RahmaAtallah, Amjed Al-Mousa, “Heart Disease Detection Using Machine LearningMajority Voting Ensemble Method”, 978-1-7281-2882-5/19/$31.00 IEEE, 2019.

[9] Gautam Chitnis,Vidhi Bhanushali, Aayush Ranade, TejasviniKhadase, Vaishnavi Pelagade, Jitendra Chavan, “A Review of Machine Learning Methodologiesfor Dental Disease Detection”, IEEE India Council International Subsections Conference (INDISCON), 2020.

[10] YedilkhanAmirgaliyev, Shahriar Shamiluulu, Azamat Serek, “Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods”, 2018.

[11] Aida Brankovic, Ali Zamani, Amin Abbosh, “Electromagnetic Based Fatty Liver Detection Using Machine Learning”, 13th European Conference on Antennas and Propagation (EuCAP), 2019.

[12] M.Ardra Meghana Simon, N. Kalyan Saradhi, P.Sahithi, R. Sai Kiran, Mrs. A. Aparna, “Liver Disease Prediction using Machine Learning”, A Journal Of Composition Theory, Volume XIV, Issue VI, JUNE 2021.

[13] Hastie T, Robert, T, Jerome F (2009). The Elements of Statistical Learning: Data mining, Inference and Prediction. Springer. 485–586.

[14] G. Shobana, K. Umamaheswari, “Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling”, 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021.

[15] Vyshali J Gogi, Dr. Vijayalakshmi M.N, “Prognosis of Liver Disease: Using Machine Learning Algorithms”, International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering - (ICRIEECE), 2018.

[16] Hartatik, Mohammad Badri Tamam, AriefSetyanto, “Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms”, 2nd International Conference on Cybernetics and Intelligent System (ICORIS), 2020.

[17] SateeshAmbesange, VijayalaxmiA, Rashmi Uppin, Shruthi Patil, Vilaskumar Patil, “Optimizing Liver disease prediction with Random Forest by

various Data balancing Techniques", IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), 2020.

[18] R. Kalaiselvi, K. Meena, V. Vanitha, "Liver Disease Prediction Using Machine Learning Algorithms", International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Oct 2021.