International Journal of Research in Advent Technology, Vol.7, No.3, March 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Liver Disease Prediction and Diagnosis Expert System using Data Mining Techniques

Dr. N. V. Ramana Murthy[1], S. Shruti[2], V. Vinay Bhargav[3], S. Anil Kumar[4]
Associate Professor[1], UG Students[2,3,4]
Department of Computer Science and Engineering[1,2,3,4]
Gayatri Vidya Parishad College for Degree and PG Courses, Rushikonda
drnvrmurty@gvpcdpgc.edu.in[1], shrutisanjeevi27@gmail.com[2],vvinaybhargav@gmail.com[3], anilkumarsenna143@gmail.com[4]

Abstract-In recent days there is increase in deaths due to liver disorder problems. Liver is the largest internal organ and gland in human body. The liver functions involve in Digestion, Metabolism, Immunity and supply nutrients in the body. projection of liver disorders at starting stages can lower the risks and diagnosis at early stage can definitely heal. the main concern of this project is to design and develop a medical diagnosis expert system which helps the physicians in decision making through collected data of liver disorders by using understanding criterions. classification algorithms like Decision Trees(J48), Naive Bayes, Random Forest and Multilayer Perceptron are used to sort and contrast the success and rate of correction of the data. It helps in implementation of classification models in terms of correctness and reduce the evaluation time is required to develop models that can grasp faster with better conception of models. a comparative investigations of data classification precision using data of liver disorders are represented. Comparisons based on performances between classifier algorithms is considered quantitatively.

Keywords-classification algorithms, liver disorders, ILPD, WEKA, Diagnosis prototype, Random Forest, Decision Tree, Multilayer Perceptron, Naïve Bayes.

1. INTRODUCTION

According to recent works, India reached 2.95% of total deaths caused due to liver diseases. cirrhosis is the 14th major cause of deaths in the world. India holds 63rd place in world rank that causes death due to liver diseases[6]. Jaundice is a common symptom of liver diseases. Jaundice is formed when there is massive amount of bilirubin in your system. bilirubin is a yellow coloured pigment created by the break-down of red blood cells in liver. Usually, the liver cleanse bilirubin along with old red blood cells. Jaundice are categorized by where they develop within the liver's process of taking in and filtering out bilirubin, hemolytic or Pre-hepatic jaundice is when the process takes place before the liver. In this pre-hepatic jaundice, the unconjugated bilirubin is present. if the process takes place in liver then it is called Hepatic jaundice. In this, the combination of both unconjugated and conjugated bilirubin is present. If the process takes after the liver then it is called Post-hepatic jaundice, in this the conjugated bilirubin is present.

1.1. Causes of liver disorder

The causes of liver disorders are categorized into congenital disorder and acquired disorder. People suffering from liver diseases can be only from either of these ways. Congenital liver deficiency is a liver disorders that are present from the birth. It is also called as genetic liver disease as it is genes from parents. These defects are usually rare. This disorder influences the flow of bile where it blocks the bile duct. Bile is a fluid produced in liver which helps in digestion of fats. Bile duct carries bile from liver to gall bladder and small intestine. Genetic liver defects have liver conditions depending upon the majority bilirubin present in the system. Gilbert's Syndrome is a liver condition occurred when increase of unconjugated bilirubin in blood while Rotor's Syndrome is a liver condition occurred when increase of both conjugated bilirubin and unconjugated bilirubin in blood but majority is conjugated bilirubin.

Acquired liver deficiency are liver disorders that are developed after birth which are not inherited from genetic. The acquired liver defects are classified into hemolytic jaundice, hepatic jaundice and obstructive jaundice. Haemolytic jaundice is a condition where increase in blood rate in process of breaking down the red blood cells and releasing of haemoglobin. The major causes of hemolytic jaundice are Malaria, Sickle Cell Anemia, Spherocytosis, Thalassemia. Hepatic jaundice is a condition observed when liver tissue is damaged. It cannot filter bilirubin from blood functionally and this results in increase of bilirubin in liver. The major causes are Liver Cirrhosis, Liver Cancer, Alcoholic Hepatitis, Viral Hepatitis, Primary Biliary Cirrhosis. Obstructive jaundice is observed when bilirubin is not filtered properly into bile ducts due to blockage. The common causes are Pancreatic Cancer, Gallstones, Bile Duct Cancer, Pancreatitis[5].

1.2. Symptoms and Signs

Symptoms are internal indication of disease which one can sense. There are numerous symptoms for liver disorder where few external symptoms are fatigue, itchy skin, fever, weakness, loss of appetite, muscles or joint pain, abdominal pain. Signs are actual indication of disease observed by the physician such as jaundice, pallor, hepatomegaly.

1.3. Hazards Aspects

Factors which can increase the jeopardy of liver disease includes heavy alcohol consumption, injection of drugs with dirty needles, obesity, diabetes, tattoos, exposure to toxins, blood transfusion.

Hepatitis C, hepatitis B, fatty liver disease, primary biliary cirrhosis, Wilson's disease, hemochromatosis and other inherited liver diseases, there is a high risk for liver cirrhosis (which is end stage of liver) if any of conditions are present.

1.4. Preventions

Vaccination, awareness and screening are ways to prevent from liver diseases. Vaccinations are available for hepatitis A and B and get tested for hepatitis C virus and seek treatment if necessary. Maintain healthy diet and regular exercise. Moderate intake of alcohol and treatment as prescribed. Avoid contact with other people blood and body fluids. Avoid blend of alcohol and medicines. Avoid overdose of medicines.

2. RESEARCH WORKS

In recent times, there are some authors and researchers worked on the diagnosis of liver disorders. Veena G. S, D Sneha, Deepti Basavaraju and Tripti Tanvi (2018) initiated "Effective Analysis and Diagnosis of Liver Disorder". In this study, researchers have considered dissimilar algorithms for developing a diagnosis prototype and evaluate execution in terms of accuracy and correctness. They performed different techniques on classification algorithms for betterment of results when the initial results were inadequate. The ultimate stage has preferable and improvised accuracy results. They produce the classification model for predicting and diagnosis of liver disease at initial stages[1].

H. Jin et.al, (2014) initiated "Decision Factors on Effective Liver Patient Data Prediction". The study is conducted on the dataset of 'Indian Liver Patient data' (ILPD). The WEKA tool is used to approve the model, as it is essential module to implement the data mining. In this study, various classification algorithms such as random forest, K Nearest Neighbors, decision trees, logistic, Naïve Bayes and Multilayer Perceptron. These algorithms are contrasted based on several evaluation criteria like accuracy, sensitivity, specificity and recall. The Logistic and Random Forest algorithms has good outcomes of accuracy and recall[2].

Tapas Ranjan Baitharua, Subhendu Kumar Pani (2016) initiated "Analysis of Data Mining Techniques for Healthcare Decision Support System Using Liver Disorder Dataset". In this Study, 6 approved algorithms like VFI, margin curve, KNN, ZeroR, decision tree and multilayer perceptron are considered to evaluate the performance. This research is used to predict the liver diseases by classification algorithm. Based on the accuracy and evaluation values, classification is done[3].

P. Saxena et.al, (2013) conducted "Analysis of different Clustering Algorithms of Data Mining in the field of Health Informatics". The clustering techniques of data mining is evaluated on ILPD dataset. The WEKA tool is used to implement the clustering algorithms such as DBSCAN, Hierarchical clustering, COBWEB and K-means clustering algorithms. The K-means clustering algorithm is most efficient when contrasted with other algorithms in the conducted experiment[4].

3. METHODOLOGY

A. Liver Dataset

The dataset preferred in this study is "Indian liver patient dataset" from University of California in Irvine (UCI) machine learning repository. The ILPD contains 11 distinct attributes of total 583 patient records where 416 are liver patients and 167 are healthy patients. This dataset holds records of 142 female and 441 males. The last attribute "selector field" was grouped into 1 and 2 by specialist based on presence and absence of liver disease[7].

Table 1. Data set attributes

S.NO	ATTRIBUTES	ATTRIBUTE INFORMATION	ATTRIBUTE TYPE
1	Age	Age of patient	Numeric
2	Gender	Gender of patient	Nominal
3	TB	Total bilirubin	Numeric
4	DB	Direct bilirubin	Numeric
5	ALKPHOS	Alkaline phosphatase	Numeric
6	SGPT	Alamine aminotransf	Numeric

		erase	
7	SGOT	Aspartate aminotransferase	Numeric
8	TP	Total proteins	Numeric
9	ALB	Albumin	Numeric
10	A/G RATIO	Albumin and globulin ratio	Numeric
11	SELECTOR FIELD	Labelled by experts	Binominal

B. Classification

In this study, the main desire of data mining is predicting and represent. Prediction is conducted on the existing values to guess the unknown patterns and representation is focuses on discovering patterns for the data interpreted by user.

Classification algorithm main objective is to predict the patterns by analysing existing dataset. It is a supervised learning approach which analyses the test dataset and produces an inferred function. Classification algorithms approach is often uses decision tree or neural network. learning and classification involves in classification process of data. In investigation process, the test data is analysed by classification algorithms. In classifying, the experiment data is used to evaluate accuracy of rules. The rules to new input are applied if precision is satisfied. The classification algorithms use pre-classified instance for proper distinction.

a) Data Pre-processing

Pre-processing of data is foremost measure to interpret every machine learning complication. In order to practise machine learning algorithms, datasets need to cleanse and alter. The frequently used pre-processing technique is to substitute the missing values of the attributes. Even though, the techniques appear in straight-forward, when dealing with dataset things get complicated and produces rare challenges. There are few missing values in dataset which are replaced to train the algorithms.

b) Classification Algorithms

i. Decision trees(J48)

J48 is an essential decision tree classifier. Decision tree is a popular machine learning algorithm implemented for classification and prediction. Decision tree classifiers generally divides into different paths, when the rule is satisfied then it proceeds along the path. In this decision tree, each node represents attribute, each branch(link) represents a decision rule and each leaf node (terminal node) represents outcome. The attributes which helps in forecast the other values of dependent variables are well known as independent variables.

ii. Random Forest

Random forests also known as random decision forests which are used to construct the predictive models for both classification and regression by desired set of techniques. Ensemble techniques use several learning models to acquire preferable outcomes. For better outcome, random forest generates forest by random uncorrelated decision trees. Each tree prefers a particular class. The class which is mostly preferred is picked by the forest.

iii. Naïve-Bayes

Naïve Bayes classifiers are group of classification algorithms based on Bayes theorem. The common principle is shared is every set of attributes existing classifier is independent.

Bayes theorem,

$$P(A|B) = P(B|A)\ P(A) / P(B)$$

Using Bayes theorem, we find the probability of A, given that B exists where B is evidence and A is hypothesis. Here, we assume that attributes are independent has presence of one attribute does not affect the other.

iv. Multilayer Perceptron

A Multilayer Perceptron (MLP) is a feed forward ANN. It consists of minimum of three layers of neurons, where neurons are known as perceptrons. MLP composed of more than one perceptron. The input signal is organised to receive the signal and the output layer is organised to make decisions or prediction concerning the input. MLPs with one hidden layer are capable of approaching sustained function. A supervised learning technique called back propagation is used for training the network where training involves in adjusting the weights in order to minimize the errors. MLPs are typically used in supervised learning problems where there is training set of input-output pairs and network must learn to model the correlation between them. In forward pass, the signal proceed from input layer through hidden layers to the output layer and outcome is measured against truth table. In backward pass, chain rule of calculus, partial derivatives of error function with respective to biases through MLP.

4. IMPLEMENTATION AND RESULTS

A. Data mining module

1. Weka

WEKA (www.cs.waikato.ac.nz/ml/weka/), is an open source data mining tool. It is developed by the University of Waikato in New Zealand that

implements data mining algorithms using the JAVA language. WEKA is used to progress the machine learning (ML) approaches and its application to real-world data mining issues. It is a group of machine learning algorithms for data mining assignment. The algorithms are executed instantly to dataset. WEKA applies algorithms for Data Pre-processing, Classification, Regression, Clustering and Visualization Tools.

2. Criteria

The algorithms are applied to the dataset ILPD and the classifier selection for prediction is based on the parameters like accuracy, precision, sensitivity and specificity.

The parameters are used are as follows

TP specifies the rate of true positive
TN specifies the rate of true negative
FP specifies the rate of false positive
FN specifies the rate of false negative

- ***Accuracy***

Accuracy is evaluated as number of all correctly classified divided by total number of datasets.

Accuracy = (TP+TN) / (TP+TN+FP+FN)

- ***Precision***

Precision is evaluated as number of correctly predicted positives divided by total number of positive predictions.

Precision = TP / (TP+FP)

- ***Sensitivity***

Sensitivity is evaluated as number of correctly predicted positives divided by the total number of positives.

Sensitivity = TP / (TP+FN)

- ***Specificity***

Specificity is evaluated as number of correctly predicted negatives divided by the total number of negatives.

Specificity = TN / (FP+FN)

Observing the efficiency of all the algorithms considered in this study. The algorithm with better outcome is preferred to generate rules in designing of expert system.

B. User Interface

The user interface is designed in static and dynamic modules. The static module contains information about liver diseases, common investigation, investigations and understanding results of liver function test. The dynamic module, the user interacts with expert system where predictions is done based on user inputs. The UI is created using html and java server pages. In dynamic module, the UI contains user input fields for symptoms and attributes in dataset. Optimization of diseases are performed based on user symptoms. The system will estimate whether the patient has a liver disease or not based on the rules in the backend. This tool can be useful for predicting and diagnosis of liver disorders at initial stage. Medical diagnosis expert system is designed in order to represent assistance for decision makers but not as substitute to them.

C. Results

Table 2. Comparisons based on Accuracy, Precision, Recall and Specificity

Classification Algorithms	Accuracy	Precision	Sensitivity (Recall)	Specificity
Naïve Bayes	85.39	94.70	76.40	95.30
Decision Tree J48	93.81	94.68	93.44	94.22
Multilayer Perceptron	98.62	98.70	98.70	98.55
Random Forest	96.21	95.20	97.70	94.58

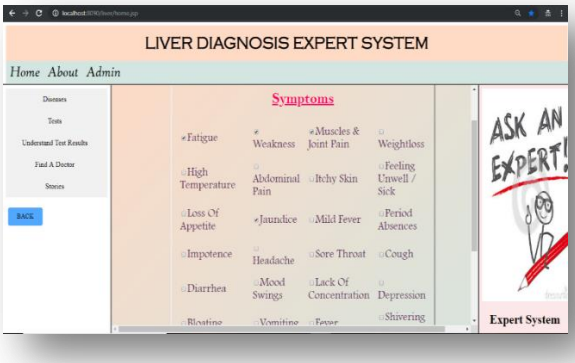

Figure 1. Symptoms Page

In symptoms page, the input is given by the user by selecting the check boxes of symptoms, which are sensed by the patients. There should be minimum of three inputs of symptoms by user to expert system to evaluate.

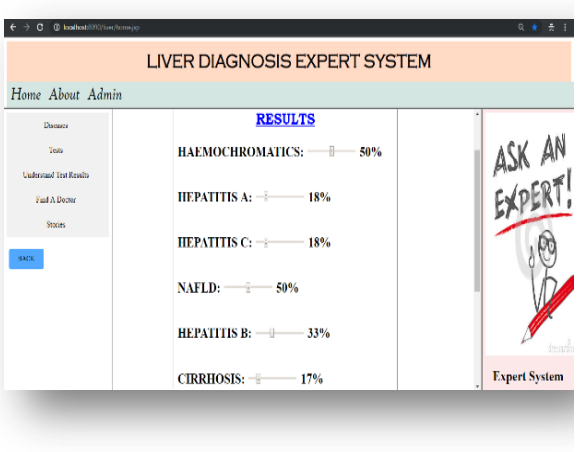

Figure 2. Results based on User Input

The results page displays the probability of liver diseases based on input symptoms given by the user.

5. CONCLUSION

In this report, various contrasting algorithms are considered for designing the diagnosis prototype and evaluating the implementation in terms of accuracy, precision and recall.

The algorithm with better outcome which can fulfil the design of expert system by rules is chosen to gain better results for prediction. Diagnosis of disorders at early stages can increase the rate of survival with a possibility of better life.

6. REFERENCES

[1] Veena G. S, D Sneha, Deepti Basavaraju and Tripti Tanvi, "Effective Analysis and Diagnosis of Liver Disorder", International Conference on Communication and Signal Processing, April 3-5, 2018, India.

[2] H. Jin et.al, "Decision Factors on Effective Liver Patient Data Prediction", International Journal of Bio-Science and Bio-Technology Vol.6, No.4, 2014.

[3] Tapas Ranjan Baitharua, Subhendu Kumar Pani, "Analysis of Data Mining Techniques for Healthcare Decision Support System Using Liver Disorder Dataset", International Conference on Computational Modelling and Security (CMS 2016).

[4] P. Saxena et.al, "Analysis of different Clustering Algorithms of Data Mining in the field of Health Informatics", International Journal of Computer & Communication Technology ISSN (PRINT): 0975 -7449, Volume-6, Issue-2, 2017

[5] Dr. M. Roopalatha, M.D (Biochemistry), Quality Head, Pinnacle Hospital, Visakhapatnam.

[6] https://www.worldlifeexpectancy.com/india-liver-disease.

[7] ILPD dataset , UCI Repository of Machine Learning https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset).