Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776735

AUTOMATED MODEL BY APPLYING MACHINE LEARNING ALGORITHMS (ML) FOR EARLY LIVER DISEASE PREDICTION

Samiksha H. Zaveri
Asst. Professor, Udhna BCA College VNSGU University,
Faculty Of Information Technology and Computer Science, Parul University, Vadodara, Gujarat, India, patel.samixa@gmail.com

Dr. Kamini Solanki
Asst. Professor, Faculty of Computer Science and Application, CHARUSAT University, Gujarat, India, solanki.kamini2311@gmail.com

Abstract
Due to excessive alcohol intake, breathing contaminated air, and eating contaminated food, liver infections have become more common swiftly. A clinical expert programme should assist the specialist in naturally anticipating the illness. The iterative growth of artificial intelligence technology would enable people to quickly recognize deadly illnesses in their early stages. It is possible to use a specialized diagnostic system at a distant location, which would be very advantageous for the medical industry. The liver is essential to maintaining good health because it helps the body get rid of toxic substances and dangerous compounds. Early detection is thus essential for diagnosis and rehabilitation. All types of AI methods are used to evaluate liver infection. It offers unique sensitivity and precision.
Keywords: Machine learning Algorithm, Liver Disease

INTRODUCTION

Liver illness, which is recognized as a severe and deadly disorder, affects people all over the globe. There is no cure if the liver is entirely damaged; the only solution is a liver transplant. Early diagnosis and treatment of the illness speed up recovery. [1]The majority of liver disease cases are difficult for medical personnel to detect at first, even if the liver's tissues are only minimally harmed. This leads to therapeutic failure. In order to prevent this, it is crucial to provide the patient the right care and save their life.

LITERATURE REVIEW

Chronic liver disease can cause a variety of digestive problems, including dry mouth, abdominal pain, internal bleeding, and constipation; skin problems, including pale skin, membranes that resemble blood vessels, and redness in the feet[2]; and abnormalities in the brain and nerves, including memory problems, numbness, and fainting. Therefore, the best measures to prevent liver disease are to see the doctor often, receive immunizations, exercise frequently, reduce your weight, and consume less soda and alcohol. [3]. According to the existing system of medical experts, the public benefits from the uncomplicated diagnosis and prognosis of liver illness that may be done via a specialized programme[4]. A number of

learning algorithms have been established via recurrent development of innovative manufacturing techniques in order to help in and improve the accuracy and quality of liver disease detection [5]. As a result, early diagnosis of liver disease is crucial for effective treatment and recovery from chronic illnesses. It is challenging to identify illnesses earlier and more accurately. [6]

METHODS

The suggested structure was developed using the Python programming language, as were the AI calculations included in this overview. The nutjob additionally, the sci-unit learn package was used to create managed classifier computations, and modules for Pandas were used to pre-handle the data. An open-source Python toolbox called NumPy does complex numerical computations on enormous amounts .A sophisticated information management tool built on the NumPy library is called Pandas

Data Preprocessing

The information gleaned from medical records is often hazy. As a result, prepping the data is essential to creating machine learning models. To ensure that the information has a comparable range of values and the components are equivalent, information preprocessing involves transforming raw information into a coherent or acceptable arrangement. In order to make the raw data more compatible with the different machine learning estimators, they were first normalized and transformed.

First, the ID column was deleted. The mean value for each variable in our dataset was used to fill in the blanks left by missing values. The information standardization method was carried out using the Standard Scalar capability, which normalizes data by scaling to unit difference due to the existence of different estimating units.

Applied classification algorithms

In this work, existing data were classified using supervised approaches, often known as classification models. The dataset had a test set (30%) and a training set (70%)From that point on, the fair preparation information was used to prepare each classifier model. Following training, the classifier classified patients using the data **.**The performance of the classifiers was assessed using the test data , and several metrics were used to compute each model's performance**.** Models are created using a variety of classifiers, which are detailed using Python v.3.5 below. [7]

Navie Bayes

This tactic uses the probabilistic Bayes' hypothesis to organize information in a controlled AI-style manner. By the Bayes theorem, all traits are considered to be independent factors. The following equation was utilised in the naive Bayesian classification**:**

a Bayesian network variation,

$P(C|X) = P(X|C)$,

is predicated on the idea that other factors may be used to forecast an event[8].

Random Forest

This tactic is yet another controlled AI method. RF used numerous decision trees built from

training data to categories the data. For each sample, this classifier employed repeated sampling to build a decision tree [9]. Some data scientists classify RF as an ensemble machine learning approach based on the decision tree model.

Support Vector Machine

A supervised machine learning technique called the support vector machine (SVM) may be used to classification and prediction problems. It enables the processing of linear and nonlinear data. To find the optimal hyperplane, the marginal distance is maximised using the SVM model. It may be used as a kernel strategy to identify the ideal sample division line. [10]

DATA SOURCE

The UCI Repository provided ILPD, the Indian Liver Patient Dataset, was used. It comes from the Andhra Pradesh region and is a typical sample of all Indians.570 occurrences (576), Eleven characteristics (Age, Gender, This dataset consists of 576 occurrences (576), Total proteins, albumin, the ratio of albumin to globulin, direct bilirubin, alkaline and alamine phosphatases, and the outcome were all measured[11]. The dataset was divided into two groups, with 100 records for group 2 and 486 records for group 1 (patients with liver disease), respectively (non-liver patients). In this study, liver illness is predicted using the Navie Bayes, Support Vector Machine, and Random Forest classifier techniques. We used a hybrid approach to attain more accuracy than current techniques.

ATTRIBUTES DESCRIPTION

Sr. No	Column Name	Data Type	Description	Range
1	Age	int64 [Numeric]	A range-containing numerical value[1]	[4 -90]
2	Gender	Object[Nominal]	both nominal values[2]	“male” or “female”
3	TB(Total Bilirubin)	Float64[Numeric]	A range-containing numerical value[3]	[0.401-74]
4	DB (Direct Bilirubin)	Float64[Numeric]	A range of numerical values [4]	[0.11-18.9]
5	Alkphos (Alkaline Phospotase	Float64[Numeric]	A ring value for a number [5]	[62-2010]
6	Sgpt (Alamine Aminotransferase)	int64[Numeric]	A range of numerical values [6]	[11-2100]
7	Sgot	Float64	(Aspartate aminotransferase) [7]	[11-4928]
8	TP (Total Proteins)	Float64	A range-containing numerical value [8]	[2.8 - 9.6]

9	ALB (Albumin)	Float64	A range of numerical values [9]	[0.8-5.4]
10	Albumin and Globulin Ratio (A/G Ratio)	Float64[Numeric]	A range-containing numerical value [10]	[0.29 - 2.87]
11	Selector Class	Numeric[1,2]	Having the class value "1" indicates having liver disease, but having the class value "2" indicates not having liver disease [11]	or 2

Table 1. Attribute Description.

PROPOSED WORK

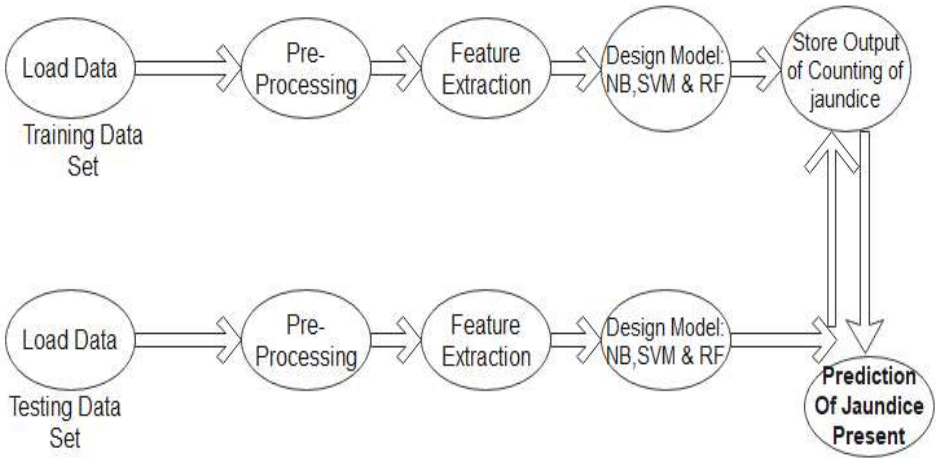

Fig. 1. Proposed work.

Algorithm Steps

(1) Start
(2) Load dataset(.csv file)
(3) Find the value of indirect Billirubin
(4) Checking the value of Jaundice using range of total billirubin, direct & indirect.(Add data in existing data frame)
(5) Apply preprocessing& labeling of data & creating feature vector (transformation).(Converts text data into Numerical on gender field)
(6) Removed rows with null value or no data. (Data cleaning/Removing null values)
(7) Splitting vector in X axis and Y axis.
  (i) Creating model by 70:30 ratio of data for NB,SVM and RF.
  (ii) Create class/category for Navie Bayes :Direct, indirect & total billirubin
  (iii) Create class/category for SVM :Direct, indirect & total billirubin
(8) Create class/category for RF :Direct, indirect& total billirubin
(9) Testing model and store result in .csv file.
(10) Load the Result file.

(11) Count majority of prediction of yes from three algorithms.
(12) Check condition: if count predication values yes >2 then
(13) Jaundice prediction=present
(14) Else Jaundice prediction=Absence
(15) Stop

CONCLUSION

Classification is the main data mining technique used in the healthcare sector for illness prediction and medical diagnosis. In this work, liver disease was predicted using a mix of Nave Bayes, Support Vector Machine (SVM), and Random Forest classification methods. Classification accuracy and confusion matrix are utilized as performance indicators in this comparison of these methods

ACKNOWLEDGMENTS

I am grateful that my mentor and god have provided me with such a significant chance in my life. Additionally, I want to thank everyone who has supported me in my work, whether directly or indirectly.

REFERENCES

[1] Predicting Liver Disease Through Machine Learning Techniques, Amshu U , Chethan Kumar M S , Mohan Lokande S, YashasPathange R ,Shraddha C , Chaya Kumari H A, (2022) International Advanced Research Journal in Science, Engineering and Technology 9(4) , ISSN (O) 2393-8021, ISSN (P) 2394-1588
[2] Dulhare UN.(2018) Prediction system for heart disease using Naive Bayes and particleswarm optimization. Biomedical Research;29(12).
[3] Subasi A, Alickovic E, Kevric J.(2017) Diagnosis of chronic kidney disease by using randomforest. CMBEBIH 2017. Springer;589–94.
[4] Shaik AB, Srinivasan S.(2019) Proceedings of International conference on innovative computing and communications. Springer;:253–60.
[5] Kotsiantis SB, Zaharakis I, Pintelas P. (2007)Supervised machine learning:a review of classification techniques. Informatica;160(1):3–24.
[6] NazmunNahar ,FerdousAra. (2018)Liver Disease Prediction By Using Different Decision Tree Techniques, International Journal Of Data Mining & Knowledge Management Process (Ijdkp) .8(2).
[7] A.Saranya, G.Seenuvasan. (2017) A Comparative Study Of Diagnosing Liver Disorder Disease Using Classification Algorithm, International Journal Of Computer Science And Mobile Computing, 6(8), .49 – 54.
[8] V.V. Ramalingam, A.Pandian, R. Ragavendran, (2018) Machine Learning Techniques On Liver Disease -A Survey, International Journal Of Engineering & Technology, 7 (4.19) .485-495.
[9] M. BanuPriya, P. Laura Juliet, P.R. Tamilselvi, (2018) Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms, International Research Journal Of Engineering And Technology, 05(01).

[10] Joel Jacob, Joseph Chakkalakal Mathew, Johns Mathew, Elizabeth Issac, (2018)Diagnosis Of Liver Disease Using Machine Learning Techniques, International Research Journal Of Engineering And Technology, 05(04) .

[11]]M. Kiran Kumar, M. Sreedevi, Y. C. A. Padmanabha Reddy, (2018) Survey On Machine Learning Algorithms For Liver Disease Diagnosis And Prediction , International Journal Of Engineering &Technology, 7 (1.8) ,.99-102.