INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 6.078
(Volume 6, Issue 3)
Available online at: www.ijariit.com

Prediction and diagnosis of liver disease

Samhitha Vishwanath
samhithavishwanath@gmail.com
Reva University, Bengaluru, Karnataka

Ankita K.
ankitakraj5@gmail.com
Reva University, Bengaluru, Karnataka

Ajay S.
ajayshan47@gmail.com
Reva University, Bengaluru, Karnataka

Annay Maity
annaymaity.cse2017@gmail.com
Reva University, Bengaluru, Karnataka

Aruna Kumara B.
arunakumara.b@reva.edu.in
Reva University, Bengaluru, Karnataka

ABSTRACT

India having a population of approx.1.3 Billion is the second most populated county in the world and each year millions are diagnosed with liver diseases. Some types of liver diseases are Jaundice, Hepatitis (A, B, C), Fatty Liver Diseases (FLD). These are caused due to consumption of alcohol, contaminated food, obesity etc. For a common man it is not practical to diagnose these diseases thus we need a system that is reliable and can predict the symptoms of the liver disease. This system predicts the liver disease using the patients' data and Naïve Bayes algorithm.

Keywords— *Naïve Bayes algorithm*

1. INTRODUCTION

The problem with liver patients is not discovered at a very early stage as it will be functioning normally even when it is partially damaged. An early diagnosis and prediction of liver disease can increase the patient's survival rate. As per World Health Organization (WHO) liver disease is the most common cause of death in India. Liver disease might affect one in five Indians. As per the recent survey, liver cirrhosis is the 14th leading cause of death in the world, and could be the 12th leading cause of death in the world by 2022.

Hepatocellular carcinoma (HCC) is the most common cause of deaths due to malignancy. Once the patient is diagnosed with the cirrhotic disease, there is very less chance of reversal of the disease. Acting against and controlling the symptoms of liver disease can mean a lot of cost expenditure for the patient. 75% of the patients are unable to afford it.

Through analysis of enzymes in the blood, liver disease can be diagnosed. Nowadays mobiles and computers are used to monitor human body conditions. Naïve Bayes classifier provides the highest performance for the liver disease dataset. The model performs both classification and prediction. A patient data set containing the enzymes data set is used in the study. The classification technique is applied (Naïve Bayes) and using this technique classification and prediction of liver disease is performed.

2. LITERATURE SURVEY

"Illustration of Random Forest and Naïve Bayes Algorithms on Indian Liver Patient Data Set" [1]

In this paper **Prof. M. Aiswarya, Prof. Swathi Srinivas and Prof. A.G. Hari Narayanan,** has focused on two different classification algorithm - **Random Forest** and **Naïve Bayes**. These classification techniques have been used to predict liver diseases. It also focuses on identifying the better algorithm with reference to its accuracy to predict the diseases. The algorithm is fed with a proper liver patient dataset. Based on the dataset provided, the algorithm is run to predict the disease with possible accuracy. The Naïve Bayes algorithm shows the outcome of belongingness of records to the corresponding classes. The highest probability class will be selected. The formula mentioned in the paper which the author has used is P (H | E) = P (E |H) x P (H) / P (E).

"A comparative study of diagnosing liver disorder disease using classification algorithm" [2]

In this paper **Research Scholar. A. Saranya**, **Asst Prof. G. Seenuvasan,** has said that data mining techniques have been very helpful in medical field. It is also used to find needed patterns with the help of decision making. Data mining plays an important role when it comes to large and huge datasets. It also focuses on how data mining helps in liver disease diagnosis and how this will help in the future. It also shows how Naïve Bayes classification algorithm is very useful in prediction of liver diseases. This paper focuses on various algorithms in classification techniques where Naïve Bayes algorithm gives the most accuracy and precision. It also focuses on C4.5 algorithm and specifies a proper of view of the advantages and disadvantages of each algorithms. It specifies that Naïve Bayes is fast and easy and requires only a small training data to give an approximation of necessary parameters for classification. The whole paper summarizes to show how classification technique is a huge help in the

medical field and would be a great advantage in the future with few more adjustments and developments. It is one of the vital things for analyzing huge patients records and datasets.

"Liver Disease Prediction Using Bayesian Classification" [3]
In this paper **Asst. Prof. S. Dhamodharan**, has explained about the increase of liver disease patients due to various reasons. There are various diseases caused in the liver and to diagnose these diseases, the author has mentioned 2 classification techniques Naïve Bayes algorithm and FT Tree. For the Naïve Bayes classifiers to work, many attributes need to be provided and based on the attributes in various patients, the algorithm gives an approximately accurate prediction. There might be a total of 12 attributes. Here, it explains that the main motive is to create a web application to predict liver disease. The algorithm is then selected using each supervised training sets using the help of various tools. Here the tool used is WEKA tool. In this paper 29 training datasets are used with 12 attributes. These are provided in the tool in both the algorithms. It is processed and run to check the accuracy of both the techniques. At the end, it is found that Naïve Bayes algorithm has better accuracy. To summarize everything, the paper has showed that Naïve Bayes is one of the best algorithm in the classification technique for liver disease prediction. It is showed that Naïve Bayes plays a vital role for the accurate prediction of liver disease.

"Application of Machine Learning in Disease Prediction" [4]
In this paper we apply different classification algorithms, each with its own advantage on three separate databases of disease available in UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modelling using the p-value test.

3. ALGORITHM

Naïve Bayes' Algorithm

Naive Bayes classifier calculates the probability of an event in the following steps:

Step 1: Calculate the prior probability for given class labels.
Step 2: Find Likelihood probability with each attribute for each class.
Step 3: Put these values in Bayes Formula and calculate posterior probability.
Step 4: See which class has a higher probability, given the input belongs to the higher probability class.

Implementation algorithm

Step 1: Data analysis.
Step 2: Read the training and test data from the liver patients csv file
Step 3: Describe statistical information about numerical columns in the dataset.
Step 4: Check for null values.
Step 5: Using visualization methods (join plots and scatter plots) we find the relationship between the features.
Direct Bilirubin & Total Bilirubin Aspartate Aminotransferase & Alamine Aminotransferase Total Proteins & Albumin Albumin and Globulin Ratio & Albumin
Step 6: Import the machine learning modules.
Step 7: Implement naïve Bayes algorithm.
Step 8: Predict output

3.1 Equations

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

P(h) is the probability of statement h is true.
P(D) is the probability of data.
P(h|D) is the probability of statement h given the data D.
P(D|h) is the probability of data D given that statement h was true.

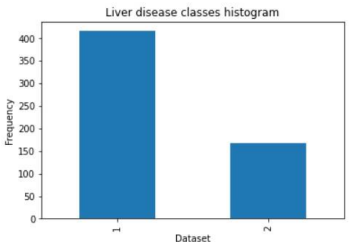
4. RESULT

When we run the program, we get the attributes of the csv file of the patients. We can see that all the attributes are not-null, this helps us to identify any null values or missing values. this also shows us the data type of each attributes.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
Age                         583 non-null int64
Gender                      583 non-null object
Total_Bilirubin             583 non-null float64
Direct_Bilirubin            583 non-null float64
Alkaline_Phosphotase        583 non-null int64
Alamine_Aminotransferase    583 non-null int64
Aspartate_Aminotransferase  583 non-null int64
Total_Protiens              583 non-null float64
Albumin                     583 non-null float64
Albumin_and_Globulin_Ratio  579 non-null float64
Dataset                     583 non-null int64
dtypes: float64(5), int64(5), object(1)
memory usage: 50.2+ KB
```

```
count_classes = pd.value_counts(data['Dataset'], sort = True).sort_index()
count_classes.plot(kind = 'bar')
plt.title("Liver disease classes histogram")
plt.xlabel("Dataset")
plt.ylabel("Frequency")
```

```
Out[5]: Text(0, 0.5, 'Frequency')
```



The above represents a graph that shows the liver disease patients (1) and the non-liver disease patients (2). The data is collected from the CSV file.

System Requirements:

(a) Intel Core i5 processor 8GB RAM
(b) Jupyter editor
(c) Machine Learning tool kit

Flow Chart:

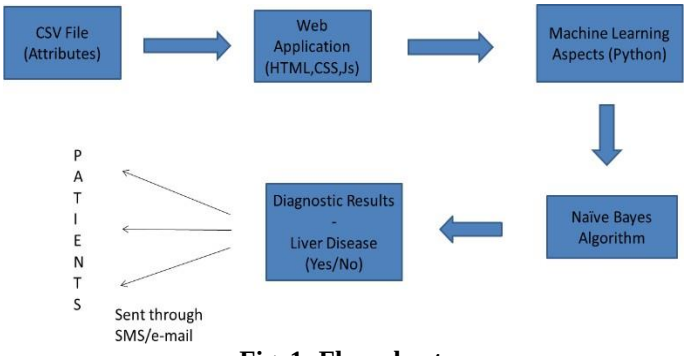


Fig. 1: Flow chart

5. FUTURE ENHANCEMENTS

This system can be further enhanced by applying multiple other classification techniques to improve accuracy. Another

future application is that this same classification and prediction techniques can be used to other diseases such as diabetes and heart diseases.

6. CONCLUSION

In this study Naïve Byes algorithm is used for prediction of liver disease in patients. During literature survey it was found that This algorithm is the most effective in terms of reliability and accuracy. And to make the procedure more seamless and faster for the patient the result of the diagnosis is sent through electronic mail.

7. REFERENCES

[1] Prof. M. Aiswarya, Prof. Swathi Srinivas, Prof. A.G. Hari Narayanan – "Illustration of Random Forest and Naïve Bayes Algorithms on Indian Liver Patient Data Set" - International Journal of Computer Science and Mobile Computing, Vol.6 Issue.8, August- 2017, pg. 49-54.

[2] Research Scholar. A.Saranya, Asst Prof. G. Seenuvasan - "A comparative study of diagnosing liver disorder disease using classification algorithm" - International Journal of Pure and Applied Mathematics, Volume 119 No. 10, 2018, 585-595.

[3] Asst. Prof. S. Dhamodharan - "Liver Disease Prediction Using Bayesian Classification" - Special Issue, 4th National Conference on Advanced Computing, Applications & Technologies, May 2014.

[4] Prof. Bendati Venkata Ramana, Prof. M. Surendra Prasad Babu, Prof. N. B. Venkateswarlu - "A critical Study of Selected Classification Algorithms for Liver Disease Diagnosis" - International Journal of Database Management System(IJDMS),Vol. 3, N0.2,2011.

[5] P S Mukherjee - "Mode of presentation of chronic liver disease in India" - 26, Oct 2017.

[6] Rong-Ho Lin - "A model for liver diagnosis. Machine Learning in Medicine" – 2017, 47:53.

[7] Times of India - "Liver disease the next major lifestyle disease of India after diabetes and BP" - 2017-11-April.

[8] Prof. Shalpa Rani Ghosh and Prof. Sajjid Waheed - "Analysis of liver disease diagnoses" - Journal of Science, Technology and Environment informatics, 05(01), 361-370, 2017.

