Survey on machine learning algorithms for liver disease diagnosis and prediction

M. Kiran Kumar *, M. Sreedevi, Y. C. A. Padmanabha Reddy

Assistant Professor, Department of CSE, Madanapalle Institute of Technology & Science

*Corresponding author E-mail:

Abstract

Machine learning plays a vital role in health care industry. It is very important in Computer Aided Diagnosis. Computer Aided Diagnosis is a quickly developing dynamic region of research in medicinal industry. The current specialists in machine learning guarantee the enhanced precision of discernment and analysis of diseases. The computers are empowered to think by creating knowledge by learning. This procedure enables the computers to self-learn individually without being explicitly programmed by the programmer .There are numerous sorts of Machine Learning Techniques and which are utilized to classify the data sets. They are Supervised, Unsupervised and Semi-Supervised, Reinforcement, deep learning algorithms. The principle point of this paper is to give comparative analysis of supervised learning algorithms in medicinal area and few of the techniques utilized as a part of liver disease prediction.

Keywords: Liver Disease; Medical Data Mining; Supervised Learning; Machine Learning Techniques.

1. Introduction

The liver is a vast, substantial organ in the human body. Weighing around 3 pounds. The liver contains two vast segments, called the right and the left projections. The gallbladder sits under the liver, alongside parts of the pancreas and digestive organs. The liver and these organs work together to process, ingest, and process sustenance. The liver's main role is to filter the harmful substances in the blood originating from the digestive system, before passing it to the rest of the body [1]. Liver damage is the one of the top deadliest disease in the world. The main causes of liver damage are Fatty liver, Liver Fibrosis, Cirrhosis, hepatitis and infections [2]. Fig 1. Shows the stages of liver damage, in the first stage healthy liver will become fatty liver due to accumulation of cholesterol and triglycerides, after few months to years fatty liver will becomes liver fibrosis, later it leads to final stage of liver damage known as cirrhosis.

In the early stages of the liver disease, it is very difficult to identify even though liver tissue has been damage moderately, it originates many medical experts repeatedly fail to diagnose the disease. This can distort to wrong medicine and treatment, so early detection is very important and necessary to save the patient [5].
Unsupervised learning is particularly known as clustering. Clustering is ever present and a wealth of clustering has been developed to solve different problems in different specific fields. However, there is no clustering algorithm that can be universally used to solve all problems. “It has been very difficult to develop a unified framework for reasoning about it (clustering) at a technical level and profoundly diverse approaches to clustering” [9]. According to AK Jain [8] clustering methods are classified into five categories partitioning, hierarchical, Density based, Grid based, Model based methods.

c) Semi Supervised Learning
Semi supervised learning is another type of machine learning. It falls between supervised learning (training data with label) and unsupervised learning (training data without label). These algorithms perform very well when we have less number of labeled data and large amount of unlabeled data [10].

d) Reinforcement Learning
Reinforcement learning is another type of Machine Learning algorithms which permits software agents and machines to automatically define the ideal behavior within the specific context, in order to maximize the performance. Simple reward feedback is required for the agent to learn its behavior. This process is known as reinforcement signal [11].

2. Background

Machine learning algorithms are very helpful in providing vital statistics, real-time data, and advanced analytics in terms of the patient’s disease, lab test results, blood pressure, family history, clinical trial data, and more to doctors. Now a days Machine learning algorithms are very useful for extracting and examining the medical data in order to build certain prediction models to rise the accuracy of diagnosis in any specific disease. However, only few works in machine learning investigate liver disorders, although this disease is aggressively increasing and becoming one of the most fatal diseases in some countries [12].

Omar S. Soliman, Eman Abo Elhand used two algorithms, one is Particle Swarm Optimization algorithm and another algorithm is Least Squares Support Vector Machine (LS-SVM) to propose a hybrid classification model for HCV diagnosis. Authors used Principle Component Analysis algorithm for extraction of feature vectors. Modified-PSO Algorithm is used to search for the optimal values of LS-SVM parameters. The proposed model was implemented and evaluated on the target HCV data set from UCI repository databases. From the experimental results the proposed system obtained highest accuracy than the other systems [13].

Moloud Abdar, Mariam Zomorodi-Moghadam, Resul Das, I-Hsien Ting, proposed computer aided Diagnostic method by using novel tree based algorithms, which are C5.0 algorithm and Chi-square Automatic Interaction Detector (CHAID) algorithm for liver disease prediction. in this proposed method authors used C5.0 algorithm via Boosting technique to achieve the highest accuracy as well as the production of rules on liver disease dataset [14].

Sadiyah Noor Novita Alfisarin, Teddy Mantoro [15] were applied three techniques, which are Decision Tree, Naive Bayes, and NBTree algorithms for diagnosis of liver disease. They have implemented classification model and obtained highest accuracy by using NB Tree algorithm. Authors concluded that the Naive Bayes algorithm gives the fastest computation time followed by Decision Tree and NB Tree algorithm and also proved that number of classification rule of NB Tree algorithm is simpler than the number of classification rule produced by Decision Tree algorithm.

Bendi Venkata Ramana, proposed five classification algorithms for liver disease diagnosis. Authors applied Naive Bayes classification, C 4.5 Decision Tree, Back Propagation, K-Nearest Neighbor and Support Vector Machine algorithms on two different type of liver Data sets, which are BUPA liver disorder data and India Liver Patient Data (ILPD). in this paper the above algorithms are considered for evaluating their classification performance in terms of Accuracy, Precision, Sensitivity and Specificity in classifying liver patient’s dataset.

![Fig. 2: Supervised Learning Process.](image-url)
Finally they obtained highest accuracy by using K-Nearest Neighbor and Support Vector Machine algorithms [16].

Sumedh Sontakke, Jay Lohokare, Reshul Dani, proposed two methods in order to classify the chronic liver disease, one method is a symptomatic approach to diagnosis, and second one involves a genetic approach to the diagnosis. Proposed approach is the application of Artificial Neural Networks and Multi-Layer Perceptron to Micro-Array Analysis. They used these two methods to improve the efficiency of two algorithms Back Propagation and Support Vector Machine (SVM) to classify the liver disease. Authors achieved highest accuracy by using Back-Propagation algorithm [17].

<table>
<thead>
<tr>
<th>Table 1: Comprehensive View of Machine Learning Algorithms for Liver Disease Diagnosis and Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Performance analysis of classification algorithms on early detection of Liver disease - ELSEVIER 2016[14]</td>
</tr>
<tr>
<td>Data Mining Techniques For Optimization of Liver Disease</td>
</tr>
<tr>
<td>A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis</td>
</tr>
<tr>
<td>IDMS-2011[16]</td>
</tr>
<tr>
<td>Diagnosis of Liver Diseases using Machine Learning IEEE-2017[17]</td>
</tr>
<tr>
<td>Comparison of Machine learning approaches for Prediction of advanced liver Fibrosis in Chronic Hepatitis C Patients IEEE-2016 [19]</td>
</tr>
</tbody>
</table>

3. Conclusion

This paper provides an idea of recent machine learning algorithms available for detection and diagnosis of liver disease. From the study it can be clearly observed that different supervised learning algorithms K-Nearest Neighbour and Support Vector Machine provide enhanced accuracy on detection of liver diseases.

References

[12] Rong-Ho Lin, "An Intelligent Model for Liver Disease Diagnosis," Artificial Intelligence in Medicine, 2009
https://doi.org/10.1016/j.artmed.2009.05.005
[19] “Comparison of Machine earning approaches for Prediction of advanced liver Fibrosis in Chronic Hepatitis C Patients” Somaya Hasmim, Gamal Esmat, Wafaa Elakel, Shahira Habasy, Safaa Abdel Raouf1, Mohamed Elhefnawi, Mohamed El-Adawy, Mahmoud ElHefnawi1, IEEE 2016
https://doi.org/10.1109/TCBB.2017.2690848