A Survey on machine learning techniques for the diagnosis of liver disease

Abstract— **The number of cases including liver diseases has been consequently becoming widespread in the context of excessive drinking, inhale of polluted gas, use of drugs, contaminated food and pickled food, therefore an expert system in the medical department will be able to make an automatic prediction of the situation. Since there is no doubt that machine learning technology is coming up with better and better results, it is now feasible for early detection of chronic liver disease, that way, people can get the disease diagnosed when it becomes fatal. As the population of elderly individuals grows, a system that provides more expertise in medical care and a medical expert system located in a remote area will come in handy. The liver owns and operates the great department which supports the movement of toxins away from the body. So, the early diagnosis of the disease is very much essential as disease prevents or at least delays the illness through the treatment. Different ML algorithms apply into a diagnosis including Supervised, unsupervised, semi-supervised and reinforcement learning fields, SVM, KNN, K-mean clustering, neural network, decision tree. Their accurate, precision and sensitivity differences want present. According to this paper, we surveyed and analyzed common machine learning techniques for the diagnosis and forecasting of liver diseases in the medical area. The prediction of liver disease has already been conducted by various authors. Our analysis was made based on Accuracy, Sensitivity, Precision, and Specificity.**

Keywords—Liver diagnosis, Machine learning, Expert System

I. INTRODUCTION

In accordance with the last report of the WHO published in 2017, liver disease related mortality is about 2.95%. India stands at 63rd rank in hepatic diseases among the world. [13] The liver is the biggest internal organ we have in our human body system. Liver has two lobes, namely: left lobe and right lobe. The liver weighs about 3 pounds.

The color of the traffic light [11] is a reddish-brown tone. The gallbladder is on the surface of the liver. The most significant function of the liver is to clear the deleterious and dangerous chemicals from the bloodstream before its distribution to the different areas of our body.

Liver disease also known to be on the list of awful health complications where in the world people are confronting. [14] The hepatic pathologies include, among others, fibrosis, fatty liver, cirrhosis, viral infections, excessive drink, drugs and toxins, and inherited metabolic defects.

abnormalities. If the liver is 100% fail there is no option to recover but only one solution is liver transplantation [15]. Early detection of liver disease can helpful in treatment of the disease to fast recover. The stages of liver disease are shown in the below figure.

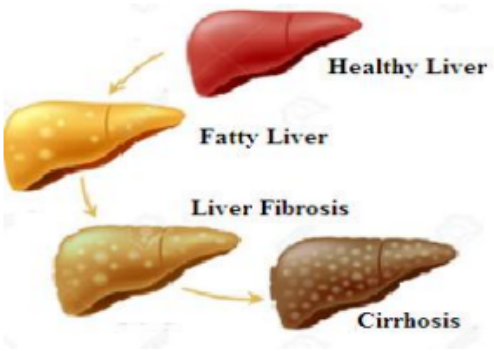

Figure 1. Liver disease stages

It can be challenging to find out in initial levels of liver disease until when the liver is significantly damaged, thus, physician systems get difficulties to point out the liver disease they have. Therefore, there is the chance of reaching the goal and guaranteeing the right implementation of the treatment and medication. Therefore, the most appropriate way to avert the prediction of this early death is by applying fixary treatment and saving the patients' life. The chronic liver disease typically comes with digestion problems; including belly pain, dry mouth, constipation and internal bleeding. Other symptoms related to skin are yellowing of the skin color, spider veins, overall redness of the feet and nerve and brain pathologies like memory problem, numbness and fainting. We can't put the blame on our liver managers for some of these diseases so some of the prevention precautions are having regular checkups by doctor, getting vaccinated , reduced soda and alcohol consumption , regular exercise and maintaining body weight. Like the medical expert system is useful to society for the diagnosis of liver disease at present also, it will be very easy to do the early detection and prediction of the disease by using the expert system. Trying and improving is the backbone of Artificial intelligence and as a result different types of machine learning algorithms are created and help in predicting liver disease quality and accuracy.

So detection of liver disease in early stages is very important and crucial because it will help in early treatment and

978-1-7281-4097-1/20/$31.00 ©2020 IEEE
2020 International Conference on Intelligent Engineering and Management (ICIEM)

recovery of the disease. And it is very difficult to detect in early stages of the disease with high accuracy.

MACHINE LEARNING

Machine learning, which is one of the Artificial Intelligence branches, enables a computer to mimic human thinking and take actions independently without needing human interference. Since then as Artificial Intelligence rapidly develops, Machine learning has made great progress in the development of diagnosis for types of diseases. As well as this, the more accurate the prediction and performance that is given by machine learning algorithms. The scope of machine learning has lately been broadly divided into supervised, unsupervised, semi-supervised, and reinforcement learning as shown in figure 2, below.

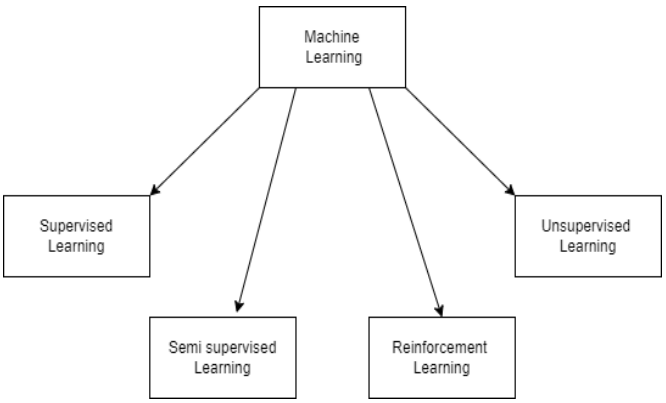

Figure 2: Different type of Machine learning

a) SUPERVISED LEARNING

Accordingly, so named supervised learning is simply one of the learning methods, which is under the guidance of a supervisor, teacher or coach. It is composed of labeled training data with patterns and makes the process of an algorithm from input to output more simple and predictable and also it occured learning and prediction process simple. There are a few supervised learning techniques, which types can be found in classification tasks such as KNN, SVM, Naïve Bayes, Neural network and regression methods like linear or polynomial, decision and Random forest. Derive stochastic forecasting from both initial concentration and final effluent.

b) UNSUPERVISED LEARNING

Clustering is a term for unsupervised learning as well. There is no ground truth set, no supervision label and the output of the unsupervised learning are the unknown data. This type of teaching method is a self-pudding learning method. Some of the supervised learning methods can be grouped under clustering techniques. Examples include K-Means clustering, SVD as well as PCA.

c) SEMI SUPERVISED LEARNING

Semi supervised learning is a type of learning method in Machine learning. The learning takes place between training data with label(SL) and training data with no label(USl). The algorithm does well for the large amount of unlabeled data and less amount of labeled data.

d) **REINFORCEMENT LEARNING**

This is a type of machine learning based on agent, action, state, reward and environment. The software agent and machine to automatically define behavior with specific context based on their reward feedback.

II. LITERATURE REVIEW

- The authors in the Bendi et al. paper [1] used two different input datasets and concluded that AP datasets are better than UCLA dataset for all algorithms considered. KNN, Backward propagation and SVM have the best performance based on their classification. The whole selected algorithm is better in the AP data set than UCLA. Naïve Bayes, C4.5, KNN, Backward propagation and SVM were found out to have accuracies of 95.07%, 96.27%, 96.93%, 97.47% & 97.07% respectively.

- They proposed a paper based on what they called Modified Rotation Forest [2], and employed two datasets for input; UCI liver dataset, and Indian liver dataset. And results have shown that MLP algorithm with random subset is better than CFS in terms of accuracy since it achieved 94.78% while the latter only had 73.07 % on UCI dataset and Indian liver dataset respectively.

- Yugal Kuma and G. Sahoo [3] offered work using a variety of classification techniques and a liver dataset from the northeastern region of Andhra Pradesh, India. The results reveal that the Decision Tree (DT) algorithm outperforms the other algorithms, with an accuracy of 98.46%.

- S.Dhamodharan [4] proposed a paper using WEKA (Waikato Environment for Knowledge and Analysis) dataset and two classification techniques (naïve Bayes and FT tree). Naïve Bayes has a higher accuracy (75.54%) than FT Tree (72.66%), indicating that it is a superior method compared to other algorithms.

- Han Ma et al. [9] in this research evaluated and demonstrated 11 distinct classifications in China Zhejiang University, College of Medicine, and concluded Bayesian network accuracy of 83%, specificity 83%, sensitivity of 0.878, and F-measure of 0.655.

- Heba Ayeldeen et al. [5] propose a paper for predicting liver fibrosis phases using decision tree techniques, employing the Cario university data set, and the results demonstrate that the decision tree classifier is 93.7% accurate.

- D. Sindhuja and R. Jemina Priyadarsini [6] review a study on the classification of liver disease. In this survey, multiple data mining classification techniques are studied, and the dataset of the AP liver outperformed the dataset of UCLA, with C4.5

achieving better results than other algorithms.

- Somaya Hashem et al. [8] published a report on the diagnosis of liver illness. This paper utilized SVM and Backpropagation methods using a dataset from the UCI Machine Repository. In conclusion, SVM has a 71% higher accuracy than Backpropagation, which has an accuracy of 73.2%.

- Joel Jacob et al. [10] suggested a study to diagnose liver illness using three different algorithms: logistic regression, K-NN, SVM, and ANN, and used the Indian Liver Patient Dataset, which consisted of ten different features for 583 individuals. Logistic regression, K-NN, SVM, and ANN achieved 73.23, 72.05, 75.04, and 92.8% accuracy, respectively..

- Sivakumar D et al. [11] developed two approaches to predict chronic liver disease. K-means and C4.5, UCI repository..

- Mehtaj Banu H [12] investigated several machine learning techniques, including supervised, unsupervised, and reinforcement learning, as well as analyzed the UCI dataset database, and concluded that KNN and SVM increased the performance and accuracy of liver disease prediction.

- Vasan Durai et al. [13] proposed a research on liver disease prediction utilizing three distinct algorithms, SVM, NB, and J48, using the UCI repository dataset, and determined that the J48 algorithm performs better in terms of feature selection, with an accuracy of 95.04%.

Table 1: Comparison table on existing machine learning technique

Sl no	Authors	Year	Disease	Machine learning algorithm	Dataset input	Remarks	Conclusion
1	Bendi Venkata Ramana et al. [1]	2011	Liver disease	Naïve Bayes, C4.5, Backward propagation, KNN and SVM	AP liver dataset and UCLA liver dataset	Naïve Bayes, C4.5 KNN, Backward propagation and SVM has 95.07, 96.27, 96.93, 97.47, & 97.07% accuracy respectively	KNN, Backward propagation and SVM are giving more better results. AP data set are better than UCLA for all the selected algorithm
2	Bendi Venkata Ramana and M.Surendra Prasad Babu [2]	2012	Liver disease	Modified Rotation Forest	UCI liver dataset and Indian dataset	MLP algorithm with random subset gives better accuracy 74.78% than NN with CFS of accuracy 73.07%	MLP algorithm with UCI liver dataset has better accuracy than NN with Indian liver dataset
3	Yugal KUMA & G. Sahoo [3]	2013	Liver disease	DT, SVM, NB and ANN	north east area of Andhra Pradesh (India) liver dataset	Decision tree(DT) has better accuracy of 98.46%	Rule based classification with DT algorithm has better accuracy
4	S.Dhamo dhar an [4]	2014	Liver cancer, Cirrhosis and Hepatitis	Naïve-Bayes, FT Tree	WEKA (Waikato Environment for Knowledge and Analysis) dataset	Naïve Bayes is 75.54% accuracy and FT Tree is 72.6624% accuracy	Naïve Bayes algorithm has better compare to other algorithms

5	Heba Ayeldeen et al. [5]	2015	Liver fibrosis	Decision tree	department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Cairo University.		decision tree classifier accuracy is 93.7%
6	D Sindhuja & R jemina Priyadarsini [6]	2016	Liver disease disorder	C4.5,Naïve Bayes, SVM, BPNN ,Regression and DT Data	AP has better dataset result than UCLA	Survey paper suggest C4.5 has better results than others	C4.5 has better accuracy result than other algorithms
7	Somaya Hashem et al [8]	2016	Liver fibrosis	PSO, GA, MReg & ADT	Egyptian national committee for control of viral hepatitis database	PSO, GA, MReg & ADT are 66.4, 69.6.69.1, & 84.4%	ADT has more accuracy result than other algorithms

						accuracy respectively	
8	Sumedh Sontakke et al	2017	Liver disease	SVM & Backpropagation	(UCI)Machine Learning Repository	SVM (accuracy 71%))& Backpropagation(accur acy 73.2%)	More accuracy result in Back propagation
9	Han ma et al	2018	Nonalcoholic fatty liver disease	Using 11 classification algorithms	First Affiliated Hospital, Zhejiang University China, College of medicine First Affiliated	Bayesian network accuracy 83%	Concluded Bayesian network has best performance than other algorithms
10	Joel Jacob et al [10]	2018	Liver disease	Logistic regression, K-NN, SVM,&ANN	Indian Liver Patient Dataset comprised of 10 different attributes of 583 patients.	Logistic regression, K NN, SVM,& ANN has 73.23, 72.05, 75.04 & 92.8% accuracy respectively	ANN has higher accuracy than others
11	Sivakumar D et al [11]	2019	Liver disease	K-means & C4.5 algorithms	UCI Repository	C4.5 algorithm has 94.36% precision.	C4.5 has better accuracy than K-means algorithms
12	Mehtaj Banu H [12]	2019	Liver disease	Supervised ,unsupervised & reinforcement	UCI repository databases.	Note: Only explaining not implementing practically	KNN and AVM has improved prediction performance accuracy
13	Vasan Durai et al [13]	2019	Liver disease	SVM,NB & J48	UCI repository	J48 algorithm has better feature selection with 95.04% accuracy	J48 algorithm is accuracy rate of 95.04%.

Table 2. Comparison table of various machine learning technique used to detect liver disease based on performance

Methods	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)	F Measure (%)
Decision Tree	98.46	95.28	95.7		
Bayesian Network	83.0	87.8	67.5		65.5
ADT*	84.4	99.0	7.0		
ANN	92.8	83.0	97.23	93.78	

J48	95.04				
BP	73.2				
SVM	71.0				

Figure 2: Performance of various machine learning technique based on their accuracy

III. CONCLUSION

This study provides an overview of previously published papers on the detection and diagnosis of liver disease using various machine learning algorithms. This survey and study definitely found and noticed that several machine learning algorithms, such as decision trees, J48, and ANN, provide higher accuracy in the identification and prediction of liver illness. And different algorithms perform differently depending on the scenario, but most significantly, the dataset and feature selection are critical for improving prediction accuracy. In addition, the study provides a survey of many types of machine learning approaches employed by different authors, with each technique having some good and bad outcomes depending on the datasets and feature selection, among other things. With this survey, we discovered that the accuracy and performance can be improved by utilising a different combination or hybrid machine learning method, and in the future, we can work on more parameters to get higher performance than the current technique.

REFERENCES

[1] Sumedh Sontakke, Jay Lohokare, Reshul Dani.”Diagnosis of Liver Diseases using Machine Learning.”International Conference on Emerging Trends & Innovation in ICT (ICEI), 2017

[2] A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain.”A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms.”INTERNATIONAL JOURNAL OF SCIENTIFIC&TECHNOLOGY RESEARCH,November 2019

[3]Ishtiaqe Hanif, Mohammad Monirujjaman Khan.”Liver Cirrhosis Prediction using Machine Learning Approaches.”

[4] Rakshith D B,Mrigank Srivastava,Ashwani Kumar ,Gururaj S P.”Liver Disease Prediction System using Machine Learning Techniques.”International Journal of Engineering Research & Technology (IJERT),Vol. 10 Issue 06, June-2021

[5]Fahad Mostafa, Easin Hasan, Morgan Williamson,Hafiz Khan.”Statistical Machine Learning Approaches to Liver Disease Prediction .”

[6]Chauhan Bhavna, Assi.Prof.Saket Swarndeep, Dr Gayatri S Pandi.”PREDICTION OF LIVER DISEASE USING MACHINE LEARNING.”International Journal of Research Publication and Reviews, Vol 3, no 5, pp 3492-3503, May 2022.

[7]Muktevi Srivenkatesh.”Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver.”International Journal of Innovative Technology and Exploring Engineering (IJITEE),Volume-9 Issue-2, December 2019.

[8]M. Banu Priya, P. Laura Juliet, P.R. Tamilselvi.”Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms.”International Research Journal of Engineering and Technology (IRJET),Volume: 05 Issue: 01 | Jan-2018 [9]Insha Arshad,Chiranjit Dutta,Tanupriya Choudhury,Abha Thakral.”Liver

Disease detection due to excessive alcoholism using Data Mining Techniques.”International Conference on Advances in Computing and Communication Engineering (ICACCE),2018

[10]Shubashini Rathina Velu,Vinayakumar Ravi,Kayalvily Tabianan.”Data mining in predicting liver patients using classification model.”Health and Technology,2022

[11]K Jeyalakshmi, R Rangaraj.”Accurate liver disease prediction system using convolutional neural network.”INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY,2021.