IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Special Issue 2, March 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

| International Conference on Recent Innovations in Engineering and Technology (ICRIET'23)|

| Sharadchandra Pawar College of Engineering, Pune, India |

|| Volume 11, Special Issue 2, March 2023 ||

Survey on Machine Learning Techniques for Diagnosis of Liver Disease

Kajal Phapale[1], Shweta Kharat[2]

Student, SPCOM, Pune, India

Sanfrancisco, California, USA

ABSTRACT: Suffering from liver disease is rapidly increasing due to excessive drinking of alcohol, inhalation of polluted gases, drugs, food contamination and food pickle packaging, so the medical expert system will help the doctor to automatically predict. With the repeated development of machine learning technology, early prediction of liver disease is possible, so people can easily diagnose the fatal disease at an early stage. This will be more useful in health department and also medical expert system can be used in remote area. The liver plays a very important role in life, supporting the removal of toxins from the body. So early prediction is very important for disease diagnosis and recovery. Various types of machine learning, supervised, unsupervised and semi-supervised, reinforcement learning for liver disease diagnosis, such as SVM, KNN, K-Mean clustering, neural network, decision tree, etc., which provides differential accuracy, precision, sensitivity. The motive of this paper is to provide an overview and comparative analysis of the entire machine learning techniques for the diagnosis and prediction of liver disease in the medical field, which are already used by different authors for the prediction of liver disease, and the analysis is based on accuracy, sensitivity, precision and specificity.

KEYWORDS: Liverdiagnosis, Machinelearning, Expert System

I. INTRODUCTION

According to the latest World Health Organization survey report published in 2017, liver disease deaths account for 2.95% of the total deaths and Indian ranks 63rd in the world [13]. The liver is the largest internal organ of our human body. The liver has two lobes, the left lobe and the right lobe. The weight of the liver is approximately 3 pounds.[11], it is a reddish-brown color. The gallbladder is located under the liver. The main important role of the liver is to remove toxic and harmful substances from the blood before distributing them to different parts of our body.

Even moderately damaged liver tissue is very difficult to identify in the early stages of liver disease, and many medical expert systems find it difficult to identify the disease in these cases. This leads to treatment and medication failure. To avoid this early prediction, it is crucial to provide the right treatment and save the patient's life. There are various symptoms of chronic liver disease such as digestive problems including abdominal pain, dry mouth, constipation and internal bleeding, dermatological problems such as yellowish skin color, spider veins, redness on the legs and abnormalities of the brain and nervous system such as memory problem, numbness and fainting. . So some of the preventative measures to prevent liver disease are regular doctor visits, vaccinations, less soda and alcohol consumption, regular exercise and weight maintenance. In addition, since the existing medical expert system for liver disease diagnosis has been useful to the society, in addition, easy disease detection and prediction can be easily done using the expert system. With the repeated improvement of artificial intelligence, various types of machine learning algorithms have been developed, which will help improve the quality and accuracy of liver disease detection or prediction. So detection of liver disease in early stages is very important and crucial as it will help in early treatment and recovery of the disease. And it is very difficult to detect in the early stages of the disease with high accuracy.

MACHINE LEARNING

Machine learning is a branch of artificial intelligence that helps a computer think like a human and can make its own decisions without human intervention. Due to the rapid development in the field of artificial intelligence, machine learning has made great progress in the diagnosis of various types of diseases. In addition, the machine learning algorithm gives us more accurate prediction and performance. Machine learning has been broadly divided into different types, which are shown in Figure 1 below.

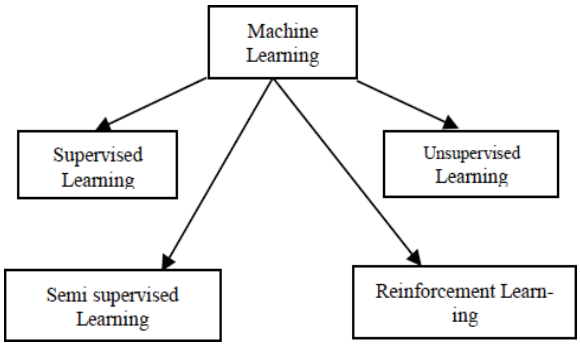

Figure1:Different types of machine learning

a) SUPERVISED EARNING

Simply put, supervised learning is a type of learning method with the help of a supervisor, teacher or instructor. Consisting of a training set of patterns associated with label data, it facilitates the algorithm from input to output, and is also easy to learn and predict. Some of the supervised learning are classifications like KNN, SVM, naive Bayes, neural network regression like linear and polynomial, decision tree and random forest. Developed prediction based on both input and output data.

In its simplest form, supervised learning refers to the process of teaching or training a computer system using labeled data. This means that the correct answer has already been assigned to certain data. In order for the supervised learning algorithm to analyze the training data (set of training examples) and provide an accurate result from the labeled data, the machine is given a new set of examples (data).

b) UNSUPERVISED LEARNING

Unsupervised learning is also known as clustering. In unsupervised learning, there is no training data set, no labels, and unknown output data. This type of learning method is like a self-guided learning method. Some of the supervised learning methods are clustering such as K-Means clustering, SVD and PCA.

Unsupervised learning is training a machine using information that is not classified or labeled and allowing the algorithm to act on that information without guidance. Here, the machine's task is to group unsorted information according to similarities, patterns, and differences without prior data training. Unlike supervised learning, no teacher is provided, meaning no training will be provided to the machine.

c) SEMI SUPERVISED LEARNING

Partially supervised learning is a type of learning method in machine learning. This learning is between labeled (SL) and unlabeled (USL) training data. This algorithm works better with large amount of unlabeled data and less amount of labeled data.

It is a technique that trains a model using both a significant amount of unlabeled data and a modest amount of labeled data. Similar to supervised learning, the goal of semi-supervised learning is to build a function that can accurately predict the output variable based on the input variables. The technique is trained on a dataset that contains both labeled and unlabeled data, unlike supervised learning. Semi-supervised learning is particularly useful when large amounts of unlabeled data are available, but it is too expensive or difficult to label them all.Simply put, a label is basically a description showing the model what it is expected to predict.

d) REINFORCEMENT LEARNING

It is a type of machine learning based on agent, action, state, reward and environment. The software agent and machine automatically define context-specific behaviors based on their reward feedback. It is about taking the appropriate action to maximize reward in a particular situation. It is used by various software and machines to find the best possible behavior or path to take in a particular situation. Reinforcement learning differs from supervised learning in that in supervised learning the training data has an answer key so the model is trained with the correct answer by itself. Perform the given task.It has to use its expertise to learn if there is no training dataset.

II. LITERATURE SURVEY

Bendi et al. [1] the authors used two different input datasets and evaluated that AP datasets have a better dataset than UCLA for all different selected algorithms. Based on their classification performance, KNN, Backward propagation and SVM give better results. The AP data set is better than UCLA for the entire selected algorithm. And I found that Naïve Bayes, C4.5, KNN, Backward propagation and SVM have accuracy of 95.07, 96.27, 96.93, 97.47 and 97.07% respectively.

Bendi et al. [2] proposed an article based on Modified Rotation Forest, used two datasets as input UCI liver dataset and Indian liver dataset. And the results show that the random subset MLP algorithm gives a better accuracy of 94.78% for the UCI dataset than CFS achieved an accuracy of 73.07% for the Indian liver dataset.

Yugal Kuma & G. Sahoo [3] proposed a work based on a different classification technique and used a dataset of livers from the northeastern region of Andhra Pradesh (India). And the results show that the decision tree (DT) algorithm is better than other algorithm and gives 98.46% accuracy.

S. Dhamodharan [4] proposed a paper based on two naive classification techniques Bayes and FT tree and used WEKA (Waikato Environment for Knowledge and Analysis) dataset. Naïve Bayes has an accuracy of 75.54% and FT Tree has an accuracy of 72.6624% and concluded that the Naïve Bayes algorithm is better compared to other algorithms.

Han Ma et al. [9] in this paper, 11 different classifiers are evaluated and demonstrated in China's Zhejiang University, College of Medicine and conclude that Bayesian network accuracy 83%, specificity 83%, sensitivity 0.878 and F-measure 0.655.

HebaAyeldeen et al. [5] proposed a work on liver fibrosis stage prediction using decision tree technique and used Cario University data set and the result shows that the accuracy of decision tree classifier is 93.7%.

D. Sindhuja and R. JeminaPriyadarsini [6] review work for classification of liver diseases. In this survey, different data mining classification techniques are studied, and the AP liver dataset used is better than the UCLA dataset, and it is concluded that C4.5 has achieved better results than other algorithms.

Somaya Hashem et al. [8] presented work for the diagnosis of liver disease. In this paper they used two algorithms, SVM & Backpropagation and used the UCI Machine Repository dataset. And it is concluded that SVM has an accuracy of 71% better result than backpropagation accuracy of 73.2%.

Joel Jacob et al. [10] proposed a liver disease diagnosis work using three different algorithms, logistic regression, K-NN, SVM and ANN, and used a dataset of Indian liver patients consisting of 10 different attributes of 583 patients. Logistic regression concluded, K-NN, SVM and ANN have accuracy of 73.23, 72.05, 75.04 and 92.8% respectively.

Sivakumar D et al. [11] proposed a work on the prediction of chronic liver disease using two different techniques K-means and C4.5. UCI repository.

MehtajBanu H [12] in this paper authors study various machine learning techniques, supervised, unsupervised & boosting and also analyze UCI data set database and concluded that KNN and SVM improved better performance and accuracy of liver disease prediction.

VasanDurai et al. [13] proposed a work based on liver disease prediction using three different techniques, SVM, NB and J48 using the UCI repository dataset and concluded that the J48 algorithm has a better performance in terms of feature selection and has an accuracy of 95.04 %.

III. CONCLUSION

This paper gives us a basic idea of a previously published paper on the detection and diagnosis of liver disease based on a different machine learning algorithm. Through this survey and study, she clearly found and observed that some machine learning algorithms such as decision tree, J48 and ANN provide better accuracy in detecting and predicting liver disease. And different algorithm has different performance based on different scenario, but most importantly, the selection of data and features is also very important to get better prediction results. And also, the article presents a survey of different types of machine learning techniques used by different authors, and each machine learning technique has some good and bad results depending on the selection of datasets and features, etc. Through this survey, we found that the accuracy and performance can be improved by using a different combination or hybrid machine learning algorithm and in the future, we can also work on more parameters to help achieve better performance than the existing technique.

REFERENCES

[1] Ramana, BendiVenkata, M. Surendra Prasad Babu,and N. B. Venkateswarlu. "A critical study of selectedclassificationalgorithmsforliver diseasediagnosis." InternationalJournalofDatabaseManagement Systems3.2(2011):101-114.

[2] Ramana, BendiVenkata, MS Prasad Babu, and N. B.Venkateswarlu."Liverclassificationusingmodifiedrotation forest." International Journal of EngineeringResearch and Development 6.1 (2012): 17-24.

[3] Kumar, Yugal, and G. Sahoo. "Prediction of differenttypes of liver diseases using rule based classificationmodel." Technology and Health Care 21, no. 5 (2013):417-432.

[4] Ayeldeen, Heba, Olfat Shaker, GhadaAyeldeen, andKhaled M. Anwar. "Prediction of liver fibrosis stagesbymachinelearningmodel:Adecisiontreeapproach."In 2015ThirdWorldConferenceonComplexSystems(WCCS),pp. 1-6. IEEE,2015.

[5] Sindhuja,D.,andR.JeminaPriyadarsini."Asurveyonclassificationtechniquesindataminingforanalyzing liver disease disorder." International JournalofComputerScienceandMobileComputing 5.5(2016):483-488.

[6] Hashem,Somaya,etal."Comparisonofmachinelearning approaches for prediction of advanced liverfibrosis in chronic hepatitis C patients." IEEE/ACMtransactionsoncomputationalbiologyandbioinformatics15.3(2017):861-868.

[7] Sontakke,S.,Lohokare,J.,&Dani,R.(2017,February). Diagnosis of liver diseases using machinelearning.In 2017InternationalConferenceonEmergingTrends&InnovationinICT(ICEI) (pp.129-133).IEEE.

[8] Ma, Han, Cheng-fu Xu, Zhe Shen, Chao-hui Yu, andYou-mingLi."Applicationofmachinelearningtechniques for clinical predictive modeling: a cross-sectional study on nonalcoholic fatty liver disease inChina."BioMedresearchinternational2018(2018).

[9] Jacob, Joel, Joseph Chakkalakal Mathew, J. Mathew,andE.Issac."Diagnosisofliverdiseaseusingmachinelearningtechniques." IntResJEngTechnol 5,no.04 (2018).

[10] SivakumarD,ManjunathVarchagall , andAmbika LGusha S "Chronic Liver Disease Prediction AnalysisBasedontheImpactofLifeQualityAttributes."(2019).InternationalJournalofRecentTechnologyand Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6S5,April2019.

[11] MehtajBanuH"LiverDiseasePredictionusingMachine-Learning Algorithms" InternationalJournalofEngineeringandAdvancedTechnology(IJEAT)ISSN:2249–8958,Volume-8Issue-6, August2019.

[12] Durai,Vasan,SuyanRamesh,andDineshKalthireddy. "Liver disease prediction using machinelearning."(2019).

[13] https://www.worldlifeexpectancy.com/life- expectancy-research

[14] D.A. Saleh F. Shebl M. Abdel-Hamid et al. "Incidenceand risk factors for hepatitis C infection in a cohort ofwomeninruralEgypt"Trans.R.Soc.Trop.Med.Hyg.vol.102pp.9219282008.https://doi.org/10.1016/j.trstmh.2008.04.011

[15] A.S.AneeshkumarandC.JothiVenkateswaran,"Estimating the Surveillance of Liver Disorder usingClassificationAlgorithms",InternationalJournalofComputer Applications (095-8887), Volume 57-No.6,November2012

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details