https://doi.org/10.46335/IJIES.2022.7.6.8 **e-ISSN: 2456-3463**
Vol. 7, No. 6, 2022, PP. 30-36

International Journal of Innovations in Engineering and Science, www.ijies.net

A Survey on Machine Learning Techniques for the Diagnosis of Liver Disease

U. Vandana [1], Dr. N. Syed Siraj Ahmed[2]

[1]PG Scholar, [2]Assitant Professor

Department of Computer Applications, Madannapalle Institute of Technology & Science, India

vandanachary6@gmail.com

***Received on**: 15 May ,2022* ***Revised on**: 14 June ,2022,* ***Published on**: 16 June,2022*

Abstract – *Suffering from liver disease has been rapidly increasing due to excessive drink of alcohol, inhale polluted gas, drugs, contamination food and packing food pickle, so the medical expert system will help a doctor to automatic prediction. With the repeated development in machine learning technology, early prediction of liver disease is possible so that people can easily diagnosis the deadly disease in the early stage. This will give more useful in the Healthcare department and also a medical expert system can be used in a remote area. The liver plays a very important role in life which supports the removal of toxins from the body. So early prediction is very important to diagnosis the disease and recovers. Different types of machine learning, Supervised, Unsupervised and Semi-Supervised, Reinforcement Learning for diagnosis of liver disease such as SVM, KNN, K-Mean clustering, neural network, Decision tree etc and give difference accuracy, precision, sensitivity. The motive of this paper is to give a survey and comparative analysis of the entire machine learning techniques for diagnosis and prediction of liver disease in the medical area, which has already been used for the prediction of liver disease by various authors and the analysis are based on Accuracy, Sensitivity, Precision, and Specificity.*

I- INTRODUCTION

As per the World health organization's latest survey report published in 2017, death due to liver disease is 2.95% of total death and Indian ranks 63rd position in the world [13]. The liver is the largest internal organ in our human body. The liver has two lobes, left lobe and right lobe. The liver weight is approximately 3 pounds[11], it's a reddish-brown color. The gallbladder is located under the liver. The main important role of the liver is to remove the toxic and harmful substances from the blood before distribution to different parts of our body. Liver disease is also considered one of the most dangerous and deadliest diseases faces in the globe. [14] The reason behind the causes of liver disease are as follows, liver fibrosis, fatty liver, liver cirrhosis, hepatitis infection excessive alcohol drink, drug and toxic and genetic abnormalities. If liver is 100% fail there is not option to recover but only one solution that is liver transplantation [15]. Early detection of liver disease can helpful in treatment of the disease to fast recover.

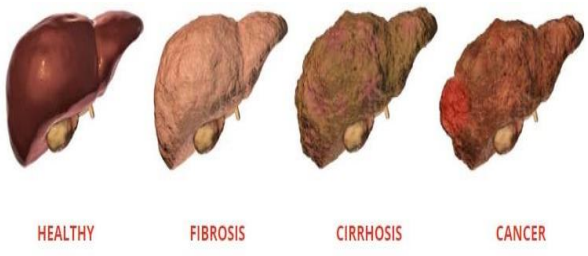


Fig 1- The stages of liver disease

https://doi.org/10.46335/IJIES.2022.7.6.8 e-ISSN: 2456-3463
Vol. 7, No. 6, 2022, PP. 30-36

International Journal of Innovations in Engineering and Science, www.ijies.net

It is very difficult to identify in early stages of liver disease even liver tissue has damaged moderately, in these case many medical expert system difficult to identify the disease. This leads to fail in treatment and medication. In order to avoid this early prediction is crucial to give proper treatment and save life of patient. There are different symptom of chronic liver disease are digestion problem including abdominal pain, dry mouth, constipation and internal bleeding, Dermatological issues like yellowish skin color, spider like veins, redness on feet and Brain and Nervous system abnormalities like memory problem, numbness and fainting. So some of the precaution to take prevention from liver disease are get regular doctor visit, get vaccinated, less soda and alcohol consumption, regular exercise and maintain weight. As per the existing system of medical expert system for diagnosis of liver disease has been useful to the society, moreover easy detection and prediction of the disease can be easy done with the use of the expert system. With the repeated improving in Artificial intelligence different types of machine learning algorithm has been developed this will help in improving the quality and accuracy of the detection or prediction of the liver disease. So detection of liver disease in early stages is very important and crucial because it will help in early treatment recovery of the disease. And it is very difficult to detect in early stages of the disease with high accuracy.

II -PROPOSED SYSTEM:

We proposed this application that can be considered a useful system it helps to reduce the limitations obtained from traditional and their existing methods.

In proposed system, we implement arandom forest algorithm for getting insights from the complex patterns in the data.This technique is robust whicj results in better accurate.

III - MACHINE LEARNING

Machine learning is a branch of Artificial Intelligence, which help the computer to think like human and can take their own decision without human intervention. Due to rapidly development in Artificial Intelligent, Machine learning has lots of advancement in diagnosis of difference types of disease. Moreover Machine learning algorithm gives us more accurate prediction and performance. Machine learning has been broadly divided into different types are shown in below figure2.

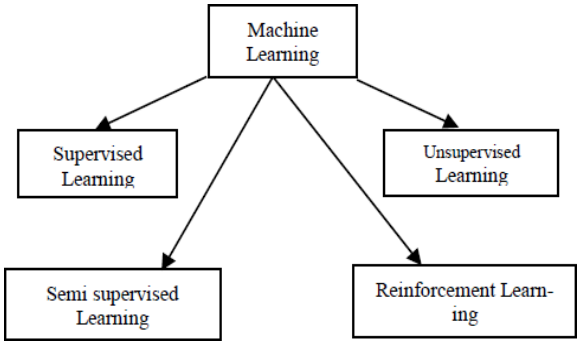


Fig 2- Different type of Machine learning

a) Supervised learning

In easy word, super vised learning is types of learning method with the help of supervisor, teacher or instructor. It consists of training set of pattern associated with label data and makes it easy for algorithm from input to output and also easy to learn and predict. Some of supervised learning are classification such as KNN, SVM, Naïve Bayes, Neural network regression as linear and polynomial, Decision tree and Random forest. Developed prediction based on both input and output data

b) Unsupervised Learning

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

c) Semi Supervised learning

Semi supervised learning is types of learning method in Machine learning, These learning is in between training data with label(SL) and training data with no label(USL).These algorithm is performing better large amount of unlabeled data and less amount of label 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

IV- LITERATURE REVIEW

- Bendi et al. [1] authors used two different input dataset and evaluate that the AP datasets has better than UCLA dataset for all the different selected algorithms. Based on performance on their

https://doi.org/10.46335/IJIES.2022.7.6.8 **e-ISSN: 2456-3463**
Vol. 7, No. 6, 2022, PP. 30-36

International Journal of Innovations in Engineering and Science, www.ijies.net

classification KNN, Backward propagation and SVM are giving better results. The AP data set is better than UCLA for the entire selected algorithm. And found out Naïve Bayes, C4.5, KNN, Backward propagation and SVM has 95.07, 96.27, 96.93, 97.47, & 97.07% accuracy respectively.

- Bendi et al. [2] proposed a paper based on Modified Rotation Forest, used two dataset as an input UCI liver dataset and Indian liver dataset. And results show that MLP algorithm with random subset gives better accuracy of 94.78% for UCI dataset than CFS achieved accuracy of 73.07% for Indian liver dataset.
- Yugal Kuma & G. Sahoo [3] proposed a paper based on different classification technique and used north east area of Andhra Pradesh (India) liver dataset. And the results shows that Decision tree(DT) algorithm has better than other algorithm and provide accuracy of98.46%.
- S. Dhamodharan [4] proposed a paper based on two classification technique naïve Bayes and FT tree and used WEKA (Waikato Environment for Knowledge and Analysis) dataset. Naïve Bayes is 75.54% accuracy and FT Tree is 72.6624% accuracy and concluded Naïve Bayes gas better algorithm compare to other algorithms.
- Han Ma et al. [9] in this paper 11 different classification are evaluated and Demonstrated in China Zhejiang University, College of medicine and concluded Bayesian network accuracy of 83%, specificity 83%, sensitivity of 0.878 and F-measure of0.655.
- Heba Ayeldeen et al. [5] propose a paper for prediction of liver fibrosis stages using decision tree technique and used Cario university data set and result shows that decision tree classifier accuracy is93.7%.
- D.Sindhuja & R. Jemina Priyadarsini [6] survey a paper for classification of liver disease. In this survey different classification techniques of data mining are study and used dataset of dataset of AP liver has better than Dataset of UCLA, and concluded C4.5 achieved better results than other
- Somaya Hashem et al. [8] presented a paper for diagnosis of liver disease. In this paper they used two algorithms, SVM & Backpropagation and used UCI machine repository dataset. And concluded SVM has accuracy 71% better result than Backpropagation accuracy 73.2%.
- Joel Jacob et al. [10] proposed a paper to diagnosis of liver disease by using three different algorithms, Logistic regression, K-NN, SVM, and ANN and used Indian Liver Patient Dataset comprised of 10 different attributes of 583 patients. And concluded Logistic regression, K- NN, SVM,& ANN has 73.23, 72.05, 75.04 & 92.8% accuracy respectively.
- Sivakumar D et al. [11] proposed a paper for prediction of chronic liver disease by using two different techniques K-means and C4.5.UCI repository.
- Mehtaj Banu H [12] in this paper authors study different machine learning technique, Supervised, unsupervised & reinforcement and also analysis UCI dataset database and concluded that KNN and SVM improved better performance and exactness of liver disease prediction.
- Vasan Durai et al. [13] proposed a paper based on liver disease prediction by using three different techniques, SVM, NB & J48 using UCI repository dataset and concluded that J48 algorithm has better performance in terms of Feature selection and has accuracy of95.04%.

Table 1: Comparison table on existing machine learning technique

Sr. 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

https://doi.org/10.46335/IJIES.2022.7.6.8 **e-ISSN: 2456-3463**
Vol. 7, No. 6, 2022, PP. 30-36

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.Dhamodhar 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, M Reg & ADT are 66.4, 69 6.69.1, & 84.4%	ADT has more accuracy result than other algorithms
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	2019	Liver disease	Supervised , unsupervised	UCI repository databases.	Note: Only explaining	KNN and AVM has

https://doi.org/10.46335/IJIES.2022.7.6.8 **e-ISSN: 2456-3463**
Vol. 7, No. 6, 2022, PP. 30-36

	[12]			& reinforcement		not implementing practically	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%.

V- INFORMATION SYSTEM

"Information system is set of **people, information technology**, and **business process** in order to achieve a business objective."

Information systems are a set of interconnected elements working together to collect, process, store, and distribute information to help coordination, visualization in an organization, analysis, and decision-making.

DECISION TREE:

Decision tree is a tree-like model that acts as a decision support tool, visually displaying Decision and their potential outcomes, consequence and cost.

BASEYESIAN NETWORK:

Bayesian network is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph.

ADT:

Abstract Data type (ADT) is a type (or class) for objects whose behaviour is defined by a set of value and a set of operations .The definition of ADT only mentions what operations are to be performed but not how these operations will be implemented. It does not specify how data will be organized in memory and what algorithms will be used for implementing the operations.

ANN:

An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks. Information that flows through the network affects the structure of the ANN because a neural network changes

J48:

C4.5 (J48) is an algorithm used to generate a decision tree developed by Ross Quinlan mentioned earlier. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier. Blood pressure (BP) is the pressure of circulating blood against the walls of blood vessels. Most of this pressure results from the heart pumping blood through the circulatory system. When used without qualification, the term "blood pressure" refers to the pressure in the large arteries

SVM:

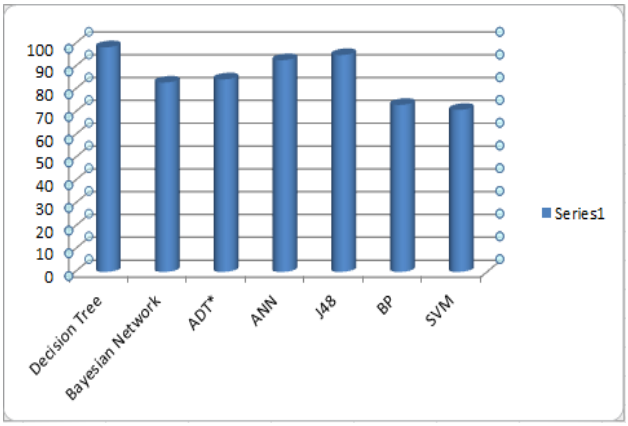
Support Vector Machine (SVM) is a relatively simple Supervised Machine Learning Algorithm used for classification and/or regression. It is more preferred for classification but is sometimes very useful for regression as well.

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	60.0	58.0
Bayesian Network	83.0	87.8	67.5	50.0	65.5
ADT*	84.4	99.0	7.0	42.0	70.5
ANN	92.8	83.0	97.23	93.78	85.0
J48	95.04	90.0	80.0	90.2	70.2
BP	73.2	89.92	57.2	70.0	40.0
SVM	71.0	75.54	80.85	78.00	85.05

https://doi.org/10.46335/IJIES.2022.7.6.8 **e-ISSN: 2456-3463**
Vol. 7, No. 6, 2022, PP. 30-36

International Journal of Innovations in Engineering and Science, www.ijies.net



VI –CONCLUSION

This paper gives us the basic idea of past published paper of detection and diagnosis of liver disease based on different machine learning algorithm. With this survey and study it has clearly find and observed that some machine learning algorithm such as Decision tree, J48 and ANN provide better accuracy on detection and prediction of liver disease. And different algorithm has different performance based on different scenario but most importantly the dataset and feature selection is also very important to get better prediction results. And also the paper presents a survey on different types of machine learning techniques used by different authors and every machine learning techniques has some good and bad outcomes depend on the datasets and features selection etc. With this survey we found out that the accuracy and performance can be improve by using different combination or hybrid machine learning algorithm and in future we can also work on more parameter which help to get better performance than the existing technique

REFERENCES

[1] *Ramana, Bendi Venkata, M. Surendra Prasad Babu, and N. B. Venkateswarlu. "A critical study of selected classification algorithms for liver disease diagnosis." International Journal of Database Management Systems 3.2 (2011):101-114.*

[2] *Ramana, Bendi Venkata, MS Prasad Babu, and N. B. Venkateswarlu. "Liver classification using modified rotation forest." International Journal of Engineering Research and Development 6.1 (2012): 17-24.*

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

[4] *Ayeldeen, Heba, Olfat Shaker, Ghada Ayeldeen, and Khaled M. Anwar. "Prediction of liver fibrosis stages by machine learning model: A decision tree approach." In 2015 Third World Conference on Complex Systems (WCCS), pp. 1-6. IEEE,2015.*

[5] *Sindhuja, D., and R. Jemina Priyadarsini. "A survey on classification techniques in data mining for analyzing liver disease disorder." International Journal of Computer Science and Mobile Computing 5.5 (2016):483-488.*

[6] *Hashem, Somaya, et al. "Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis C patients." IEEE/ACM transactions on computational biology and bioinformatics 15.3 (2017):861-868.*

[7] *Sontakke, S., Lohokare, J., & Dani, R. (2017, February). Diagnosis of liver diseases using machine learning. In 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI) (pp. 129-133).IEEE.*

[8] *Ma, Han, Cheng-fu Xu, Zhe Shen, Chao-hui Yu, and You-ming Li. "Application of machine learning techniques for clinical predictive modeling: a cross-sectional study on nonalcoholic fatty liver disease in China." BioMed research international 2018(2018).*

[9] *Jacob, Joel, Joseph Chakkalakal Mathew, J. Mathew, and E. Issac. "Diagnosis of liver disease using machine learning techniques." Int Res J Eng Technol 5, no. 04 (2018).*

[10] *Sivakumar D , Manjunath Varchagall , and Ambika L Gusha S "Chronic Liver Disease Prediction Analysis Based on the Impact of Life Quality Attributes." (2019). International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume- 7, Issue-6S5, April2019*

[11] *Mehtaj Banu H" Liver Disease Prediction using Machine-Learning Algorithms" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8Issue-6,August 2019*

[12] *Durai, Vasan, Suyan Ramesh, and Dinesh Kalthireddy. "Liver disease prediction using machine learning."(2019).https://www.worldlifeexpectancy.com/life-expectancy-research*

[13] *D.A. Saleh F. Shebl M. Abdel-Hamid et al. "Incidence and risk factors for hepatitis C infection in a cohort of women in rural Egypt"Trans. R. Soc. Trop. Med. Hyg. vol. 102 pp. 921928 2008. https://doi.org/10.1016/j.trstmh.2008.04.011*

[14] *A.S.Aneeshkumar and C.Jothi Venkateswaran, "Estimating the Surveillance of Liver Disorder using Classification Algorithms", International Journal of Computer Applications (095-8887), Volume 57-No.6, November2012*

[15] *www.techopedia.com/definition/5967/artificial-neural-network-ann*

[16] *W. Richert, L. P. Coelho, "Building Machine Learning Systems with Python", Packt Publishing Ltd., ISBN 978-1-78216-140-0*

https://doi.org/10.46335/IJIES.2022.7.6.8 **e-ISSN: 2456-3463**
Vol. 7, No. 6, 2022, PP. 30-36

[17] *J. M. Keller, M. R. Gray, J. A. Givens Jr., "A Fuzzy KNearest Neighbor Algorithm", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, No. 4, August 1985*

[18] *] S. Marsland, Machine learning: an algorithmic perspective. CRC press, 2015. [5] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine learning techniques in cognitive radios," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1136–1159, Oct. 2012. [6] https://en.wikipedia.org/wiki/Instance-based_learning*

[19] *R. S. Sutton, "Introduction: The Challenge of Reinforcement Learning", Machine Learning, 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992 [8] P. Harrington, "Machine Learning in action", Manning Publications Co., Shelter Island, New York, 2012.*