A Survey on Machine Learning Techniques For The Prediction of Liver Disease

Dr.T.S.Baskaran[1], Ms.B.Madhuvanthi[2]

[1]Associate Professor & Research Supervisor, Dept of Computer Science
[2]Dept of Computer Science
[1, 2] A. VeeriyaVandayar Memorial Sri PushpamCollege(Autonomous), Poondi-613503,Thanjavur,
Affiliated to Bharathidasan University,Tiruchirappalli-620024",TamilNadu, India

Abstract- *The liver constitutes the largest gland in the human body and performs various functions. It processes what a person eats and drinks and turns food into nutrients that the body needs to absorb. In addition, it filters harmful substances from the blood and helps fight infections. Exposure to viruses or risky chemicals can damage the liver. When this organis damaged, liver disease can improve. Liver disease refers to any condition that damages the liver and affects its function. It is a severe condition that threatens human life and needs urgent medical attention. Early diagnosis of liver diseaseis possible so that people can easily detect the deadly disease at an early stage. It is very useful in the healthcare sector and can use medical expert systems in remote areas. Predicting disease using machine learning(ML)techniques will be of interest in this study. Liver disease prediction was analyzed using classification algorithms such as Naive Bayes, SVM, Decision Tree, J48, K Star, Extra Trees and Logistic Regression. The main objective of this thesis is the early detection of liver disease.*

Keywords- Liver disease, prediction, machine learning.

I. INTRODUCTION

The liver is the largest solid organ in our human body. The liver has two lobes, the left lobe and the right lobe. The liver is located in the upper part of the gastrointestinal tract of the human body, and its weight ranges from 1400-1800 grams in males and 1200-1400 grams in females and is reddish-brown. The main role of the liver is to remove the toxic and harmful substances from the blood before distribution to different parts of our body. Viruses and alcohol use can lead the liver towards liver damage and a life-threatening situation in a person. If the liver is 100% failed there is no option to recover but the only solution is liver transplantation. There are many types of liver diseases such as hepatitis, cirrhosis, liver tumors, liver cancer, and many more. Among them liver diseases and cirrhosis as the main cause of death. Therefore, liver disease is one of the major health problems in the world. Liver diseases are usually seen in people between the ages of 40-60 and it is more common in

men. Every year, about 2 million people die worldwide from liver disease. Early detection of liver disease can be helpful in the treatment of the disease to fast recover. The stages of liver disease are shown in the below figure 1,

Figure 1 Liver disease stages

It is very difficult to identify in the early stages of liver disease even if liver tissue has been damaged moderately, in these cases many medical expert systems difficult to identify the disease. This leads to failure in treatment and medication. To avoid this early prediction is crucial to give proper treatment and save the life of the patient. There are different symptoms of chronic liver disease are digestion problems including abdominal pain, dry mouth, constipation, and internal bleeding, Dermatological issues like yellowish skin color, spider-like veins, and redness on feet and Brain and Nervous system abnormalities like memory problems, numbness, and fainting. Some of the precautions to take to prevent liver disease are regular doctor visits, getting vaccinated, less soda and alcohol consumption, regular exercise, and maintaining weight. The existing system of a medical expert system for the diagnosis of liver disease has been useful to society, moreover, easy detection and prediction of the disease can be easily done with the use of the expert system. Many diseases and disorders may be diagnosed with ML in the future. Some of these include Alzheimer, heart failure, breast cancer, and pneumonia. Medical professionals increasingly turn to machine learning (ML) algorithms to make accurate diagnoses [11]. With the repeated improvements 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 liver disease. So detection of liver disease in early stages is very important and crucial because it will help in early treatment and recovery of the disease. It is very

difficult to detect in early stages of liver disease with high accuracy.

II. LITERATURE REVIEW

A lot of work has been done related to disease prediction systems using various machine learning algorithms in health centers. Machine learning classification algorithms are widely used in many fields to solve many problems. A field like healthcare is considered a machine learning niche where machine learning can be used to address various medical decisions.

M. Banu Priya., [1] proposed performance analysis of liver disease prediction using machine learning algorithms have analyzed the liver disease using algorithms such as J48, MLP, SVM, Random Forest, and Bayesnet Classification. This algorithm gives various results based on the PSO feature selection model. It has been seen that Bayesnet and J48 Classification give better results compared to other classification algorithms.

Binish Khan., [2] strategic analysis in prediction of liver disease using different classification Algorithms have been focused on the related works of various authors on liver disease such that algorithms were implemented using the Weka tool which is a machine learning software written in Java. Various attributes that are essential in the prediction of liver disease were examined and the dataset of liver patients was also evaluated and compared various classification algorithms such as Random Forest, Logistic Regression, and Separation Algorithm to identify the best technique. The Random Forest algorithm with the highest accuracy outperformed the other algorithms.

Ain Najwa Arbain., [3] presented a comparison of data mining algorithms for liver disease prediction on imbalanced data focuses on applying a data mining algorithm to an imbalanced dataset to predict the occurrence of liver disease in patients. Before applying data mining algorithms, random sampling is done to ensure that the target class is balanced. This is important because it affects the performance of each model. K-Nearest Neighbor (k-NN) has higher accuracy rate and ROC index compared to other algorithm.

A.K.M Sazzadur Rahman., [4] presented a comparative study on liver disease prediction using supervised machine learning algorithms have been effective diagnosis system for chorionic liver infection patients utilizing six distinctive supervised machine learning classifiers. Researched all classifiers execution on patient's information parameters and the Logistic

Regression classifier gives the most elevated order exactness 75% dependent on F1 measure to predict the liver disease and NB gives the least precision 53%. From now on, the outperform classification procedure will give for the decision support system and diagnosis of chronic disease.

A.Sivasangari., [5] proposed diagnosis of liver disease using machine learning models have been liver disease is difficult to diagnose due to mild symptoms in the early stages. If it is too late, the symptoms will always appear. Hence liver related disease causes more problems to the living people and identification and recognition of the causes is very important nowadays. Therefore, an automated program is required to be developed with high accuracy and reliability for early detection of liver disease. For this purpose specific machine learning models have been developed to predict disease. In this paper, support vector machines (SVM), decision tree (DT) and random forest (RF) methods are proposed to predict liver disease with better accuracy, precision and reliability.

Md. Fazle Rabbi., [6] presented prediction of liver disorders using machine learning algorithms: a comparative study have been liver disorders can lead to various dangerous diseases, including liver cancer. Early diagnosis and treatment of patients is imperative to reduce the risk of those fatal diseases. Since diagnosis of liver disease is expensive and sophisticated, many studies have been performed using machine learning (ML) methods to classify liver disorder cases. In this paper, we have used four different ML algorithms namely Logistic Regression (LR), Decision Tree (DD), Random Forest (RF), and Extra Trees (ED) to classify the Indian Liver Patients dataset (ILPD). Pearson Correlation Coefficient Based Feature Selection (PCC-FS) is used to remove irrelevant features from the dataset. Also, a boosting algorithm (AdaBoost) is used to improve the predictive performance of those algorithms. Comparative analysis is evaluated based on precision, ROC, F-1 score, precision and recall. After comparing the experimental results, we find that increasing ET provides the highest accuracy of 92.19%.

Dr. P.Hamsagayathri., [7] presented a symptoms based disease prediction using machine learning techniques proceed in many applications, machine learning plays a vital role, such as image recognition, data mining, processing of natural languages and diagnosis of diseases. Machine learning provides potential solutions in all these fields. Discusses various techniques of Machine Learning for the diagnosis of various diseases such as heart, diabetes and liver diseases. Most models have shown excellent results because they specifically describe the characteristic. It is noted from

previous studies that K star provides 83.47 percent improved performance for liver disease identification.

Sreenivasa Rao Veeranki., [8] presented an intelligent techniques and comparative performance analysis of liver disease prediction have implemented the mostly used machine learning approaches like random forest, SVM, KNN and MLP has been used to predict the dataset of Indian Liver Patient Dataset. All of them are used to predict liver disease and thus a comparative study on their performances are created. The Support Vector Machine is performed well when it has been compared with other models.

Ketan Gupta., [9] presented a liver disease prediction using machine learning classification techniques proposed the historical and classified input of patients and output data is fed into various algorithms or classifiers for predicting the future data of patients. The algorithms used here for predicting liver patients are Logistic regression, Decision Tree, Random Forest, KNN, Gradient Boosting, Extreme Gradient Boosting, LightGB. Based on the analysis and result calculations, it was found that these algorithm has obtained good accuracy after feature selection.

K. Arumugam., [10] proposed a multiple disease prediction using machine learning algorithms evolved from database statistics and is valuable in assessing the efficacy of medical interventions. Data Visualization with Machine Learning Diabetes-related heart disease is a type of heart disease that occurs in diabetic patients. Diabetes is a chronic disease that arises when the pancreas fails to produce enough insulin or when the body fails to use the insulin produced properly. Heart disease, often called cardiovascular disease, is a group of disorders that affect the heart or blood vessels. Although there are many data-mining classification methods for predicting heart disease, there is not enough data to predict heart disease in a diabetic patient. We fine-tuned the decision tree model for optimal performance in predicting the likelihood of heart disease in diabetic patients, as it outperformed the naive Bayes and support vector machine models.

This paper provides a survey of various models based on such methodologies and techniques and reports their exploratory analysis in Table 1. Models based on supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF), J48, K Star, Extra Trees and Logistic Regression etc., are very popular among researchers and are used for the investigation of large and complex data. Systems have been applied to different clinical datasets to robotize. Many scientists, of late, are using a few machine learning algorithms, and the techniques have been applied to many clinical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help healthcare professionals and professionals diagnose liver-related diseases. Strategies for implementing the well-being of professionals in the analysis of industrial and liver-related diseases. This paper presents a review of various models based on such calculations and methods and analyzes their exhibition. Models in the light of directed learning calculations, for example, Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Naïve Bayes, Decision Trees (DT), Random Forest (RF), J48, K Star, Extra Trees, Logistic Regression etc. and have been found very well - known among scientists.

Table 1Survey analysis report

YEAR	AUTHOR	TITLE	TECHNIQUE USED	ACCURACY
2018	M. Banu Priya, P. Laura Juliet, P.R. Tamilselvi	Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms	J48, MLP, SVM, Random Forest, and Bayesnet.	95.04% Accuracy (J48)
2019	Binish Khan, Piyush Kumar Shukla, Manish Kumar Ahirwar	Strategic Analysis in Prediction of Liver Disease Using Different Classification Algorithms	Naïve Bayes, Decision Tree, Multilayer Perceptron, k-NN, AdaBoost, J48, Random Forest	Random forest with 100% accuracy
2019	Ain Najwa Arbain, B. Yushalinie Pillay Balakrishnan	A Comparison of Data Mining Algorithms for Liver Disease Prediction on Imbalanced Data	Random Forest, K-NN, Auto Neural, Logistic	K-NN with 99.766%

2019	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	Logistic Regression, K Nearest Neighbors, Decision Tree, Support Vector Machine, Naïve Bayes, and Random Forest.	Logistic Regression with 75%
2020	A.Sivasangari, Baddigam Jaya Krishna Reddy, Annamareddy Kiran, P.Ajitha	Diagnosis of Liver Disease using Machine Learning Models	Support Vector Machine, Decision Tree, and Random Forest.	Support Vector Machine with 95.18%
2020	Md. Fazle Rabbi, S. M. Mahedy Hasan, Arifa Islam Champa, Md. AsifZaman, Md. Kamrul Hasan	Prediction of Liver disorders using machine learning algorithms: A Comparative Study	Logistic Regression, Decision Tree, Random Forest, and Extra Trees	Extra Trees with 92%
2021	Dr. P.Hamsagayathri, Mr.S. Vigneshwaran	Symptoms Based Disease Prediction Using Machine Learning Techniques	Support Vector Machine, Random Forest, and K star	K Star with 83.47%
2022	Sreenivasa Rao Veeranki and Manish Varshney	Intelligent Techniques and Comparative Performance Analysis of Liver Disease Prediction	Support Vector Machine, K-Nearest Neighbour, and Multilayer Perceptron	Support Vector Machine with 74%
2022	Ketan Gupta, Nasmin Jiwani, Neda Afreen, Divyarani D	Liver Disease Prediction using Machine learning Classification Techniques	Logistic Regression, Decision Tree, K- Nearest Neighbour algorithm, Random Forest, Gradient Boosting, XGBoosting	Random Forest, Light GB, Ada booting algorithm gives better accuracy.
2023	K. Arumugam, Mohd Naved, Priyanka P. Shinde, Orlando Leiva-Chauca, Antonio Huaman-Osorio, Tatiana Gonzales-Yanac	Multiple disease prediction using Machine learning algorithms	Support Vector Machine, Navie Bayes, Decision Tree	Support Vector Machine With 91%

III. CONCLUSION AND FUTURE WORK

Summarizing different types of machine learning algorithms for the prediction of liver diseases. Elegant worked on finding the best algorithm by analyzing various machine learning algorithms and their features. Each algorithm has given different results at different levels. The predictive model of liver disease is only analyzed with partial accuracy, so more complex models are needed to increase the accuracy of predicting early liver disease. In the future, we will propose a method for early prediction of liver disease with high accuracy and low cost and complexity.

REFERENCES

[1] M. Banu Priya, P. Laura Juliet, and P. R. Tamilselvi, "Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms," *Int. Res. J. Eng. Technol.*, vol. 5, no. 1, pp. 206–211, 2018, [Online]. Available: www.irjet.net

[2] B. Khan, P. K. Shukla, and M. K. Ahirwar, "Strategic Analysis in Prediction of Liver Disease Using Different Classification Algorithms," *Int. J. Comput. Sci. Eng.*, vol. 7,no. 7, pp. 71–76, 2019, doi: 10.26438/ijcse/v7i7.7176.

[3] A. N. Arbain and B. Y. P. Balakrishnan, "A Comparison of Data Mining Algorithms for Liver Disease Prediction

on Imbalanced Data," *Int. J. Data Sci. Adv. Anal. (ISSN 2563- 4429)*, vol. 1, no. 1, pp. 1–11, 2019, [Online]. Available:
http://ijdsaa.com/index.php/welcome/article/view/2

[4] A. K. M. S. Rahman, F. M. Javed Mehedi Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, "A comparative study on liver disease prediction using supervised machine learning algorithms," *Int. J. Sci. Technol. Res.*, vol. 8, no. 11, pp. 419–422, 2019.

[5] A. Sivasangari, B. J. Krishna Reddy, A. Kiran, and P. Ajitha, "Diagnosis of liver disease using machine learning models," *Proc. 4th Int. Conf. IoT Soc. Mobile, Anal. Cloud, ISMAC 2020*, pp. 627–630, 2020, doi: 10.1109/I-SMAC49090.2020.9243375.

[6] M. F. Rabbi, S. M. Mahedy Hasan, A. I. Champa, M. Asifzaman, and M. K. Hasan, "Prediction of liver disorders using machine learning algorithms: A comparative study," *2020 2nd Int. Conf. Adv. Inf. Commun. Technol. ICAICT 2020*, pp. 111–116, 2020, doi: 10.1109/ICAICT51780.2020.9333528.

[7] P. Hamsagayathri and S. Vigneshwaran, "Symptoms based disease prediction using machine learning techniques," *Proc. 3rd Int. Conf. Intell. Commun. Technol. Virtual Mob. Networks, ICICV 2021*, no. Icicv, pp. 747–752, 2021, doi: 10.1109/ICICV50876.2021.9388603.

[8] S. Rao Veeranki and M. Varshney, "Intelligent Techniques and Comparative Performance Analysis of Liver Disease Prediction," *Int. J. Mech. Eng.*, vol. 7, no. 1, pp. 974–5823, 2022.

[9] K. Gupta, N. Jiwani, N. Afreen, and D. Divyarani, "Liver Disease Prediction using Machine learning Classification Techniques," *Proc. - 2022 IEEE 11th Int. Conf. Commun. Syst. Netw. Technol. CSNT 2022*, no. April, pp. 221–226, 2022, doi: 10.1109/CSNT54456.2022.9787574.

[10] K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, "Multiple disease prediction using Machine learning algorithms," *Mater. Today Proc.*, vol. 80, no. xxxx, pp. 3682–3685, 2023, doi: 10.1016/j.matpr.2021.07.361.

[11] S. S. Kumar and T. S. Baskaran, "Diagnosis of Medical Conditions in Patients Using Machine Learning Models," *NeuroQuantology*, vol. 20, no. 16, pp. 2289–2297, 2022, doi: 10.48047/NQ.2022.20.16.NQ880228.