ISSN: 2581-3404 (Online) IF: 5.86 (SJIF)
IJIRTM, Volume-10, Issue-3, May-2026

Performance Analysis of Liver Diseases Classification using Machine Learning Classifier

Neetiraj Singh Lodhi[1], Dr. Rajneesh Choubey[2]
M. Tech, Research Scholar[1], Associate Professor[2]
Dept. of CSE, Bansal Institute of Science and Technology, Bhopal, India[1]
Dept. of MCA, Bansal Institute of Science and Technology, Bhopal, India[2]
nrsklodhi@gmail.com[1], choubey.rajnish@gmail.com[2]

Abstract. *Early detection through supervised learning techniques applied to diverse datasets is crucial in reducing mortality rates. With ongoing advancements in medicine, a significant amount of data has been collected, particularly in healthcare, where extensive data generation occurs. This data is processed and analyzed through data mining applications for knowledge extraction. Mining algorithms effectively predict patient diseases by utilizing appropriate learning strategies. Chronic kidney disease (CKD), hepatitis, cancer, and diabetes represent major global health concerns, making their prediction a significant focus for researchers. This dissertation primarily analyzes various classification algorithms, including Support Vector Machines (SVM), K Nearest Neighbor (KNN), and Extra Tree Classifier, by comparing their performance utilizing liver patient data. The study employs different machine learning classification techniques to diagnose liver conditions at an early stage, yielding performance metrics such as accuracy and other relevant parameters.*

Keywords: Accuracy, support vector machines, Supervised machine learning, Random forest.

Introduction

Liver disease is a major global health issue that affects millions of individuals and disrupts essential metabolic functions including detoxification, protein synthesis, and digestion. Various liver conditions such as hepatitis, cirrhosis, fatty liver disease, and liver cancer can drastically impair liver function, leading to serious, sometimes life-threatening consequences. Consequently, early detection of these diseases is vital for effective treatment and prevention of severe complications. Traditional diagnostic approaches, which largely depend on clinical experience, lab tests, and imaging techniques, come with inherent limitations including constraints on time, cost, and potential for human error. This highlights a critical need for intelligent, automated diagnostic systems to enhance detection capabilities.

To understand liver failure and the importance of early detection, one must first grasp the liver's functionalities. The liver, a significant organ located in the right upper quadrant beneath the diaphragm, plays vital roles in maintaining overall health. It not only detoxifies harmful substances and metabolizes medications but also produces bile, which is essential for the absorption of fats, proteins, carbohydrates, and certain vitamins. The organ is involved in metabolizing bilirubin, generating blood-clotting factors, processing macronutrients, storing crucial vitamins and minerals, filtering blood, producing albumin, and removing damaged red blood cells. Technological advancements have positioned machine learning as a groundbreaking resource in healthcare, particularly for analyzing medical data. Machine learning algorithms can uncover intricate patterns within large datasets that might not be readily evident to healthcare professionals, thereby enabling accurate predictions regarding the likelihood of liver disease based on historical patient information. These predictive systems apply structured datasets that encompass a variety of medical attributes, including age, gender, and biochemical markers such as bilirubin and enzyme levels, to create models that estimate the probability of liver disease presence.

The prediction process typically involves several key stages: data collection, pre-processing, feature selection, model training, evaluation, and deployment. Data pre-processing is particularly important as it confronts challenges like missing values and the normalization of data. Feature selection focuses on identifying critical attributes that will enhance prediction accuracy. A variety of machine learning algorithms are utilized in liver disease prediction, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and Gradient Boosting, each contributing uniquely to the predictive modeling process.

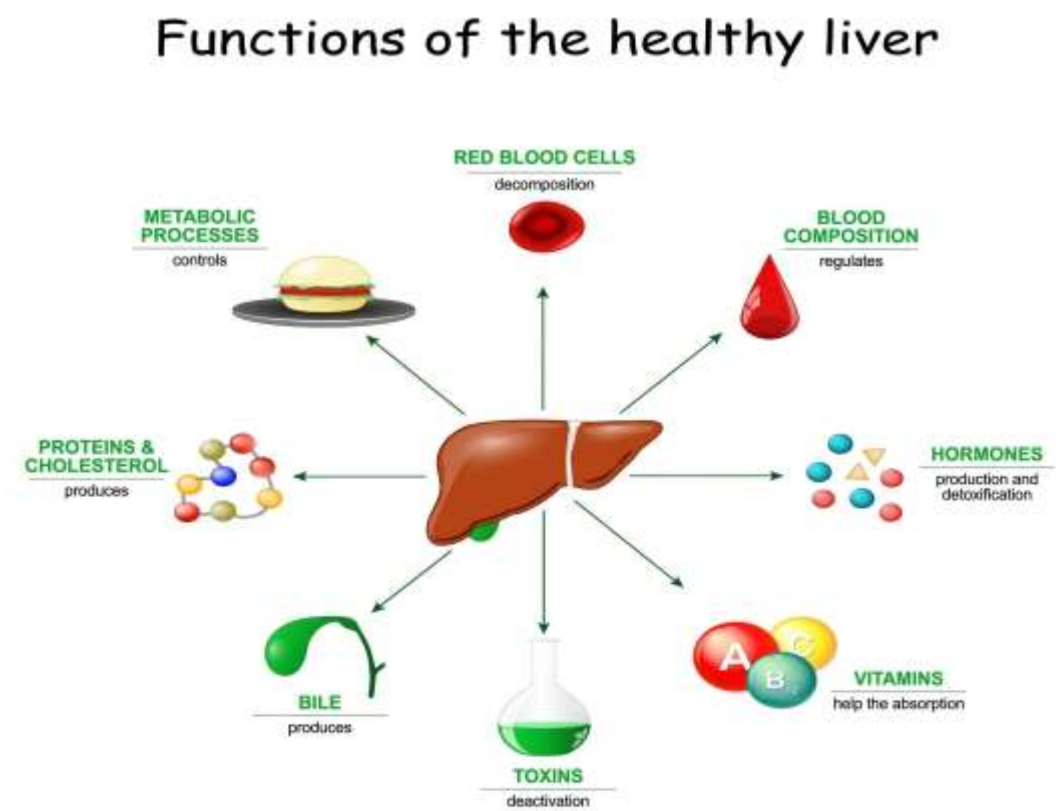

Figure 1.1: Functions performed by the liver [15].

Machine Learning

Liver disease represents a significant global health issue, particularly affecting countries like India, with disorders such as cirrhosis, hepatitis B and C, and non-alcoholic fatty liver disease (NAFLD) contributing notably to mortality rates. The progression of these diseases is often asymptomatic until advanced stages, underscoring the importance of early diagnosis for effective treatment and prevention. Traditional diagnostic methods, including biopsy, ultrasound, CT, and MRI, are commonly utilized but have limitations such as invasiveness, high costs, limited sensitivity, and reliance on expert interpretation. Recent advancements in healthcare have seen the incorporation of Machine Learning (ML) and Deep Learning (DL) technologies in the prediction and diagnosis of liver diseases. These methods leverage extensive medical datasets to reveal hidden patterns and forecast disease prognosis with high accuracy. Various ML algorithms, including Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forest, and Gradient Boosting, have been crucial in liver disease prediction. Furthermore, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated exceptional efficacy in analyzing medical images, such as ultrasound scans, to identify fatty liver disease.

Dataset Description

Indian Liver Patient Dataset is taken from the University of California, Irvine ML repository, and it accommodates 11 columns with which ten features and one target variable are used for this research are provided in Table 1 (ILPD (Indian Liver Patient Dataset) – UCI Machine Learning Repository).

Table 1: Features of ILPD Dataset.

Sr. No.	Variable name	Feature Type
1	Gender – Patient	Categorical
2	Total Bilirubin (TB)	Real number
3	Direct Bilirubin (DB)	Real number
4	Alkaline Phosphatase (Alkphos)	Integer
5	Alanine Aminotransferase (SGPT)	Integer
6	Asparatate Aminotransferase (SGOT)	Integer
7	Total Proteins (TP)	Real number
8	Albumin(ALB)	Real number
9	Albumin and Globulin Ratio (A/G)	Real number
10	Classes used for the dataset	Categorical

The ILPD dataset encompasses data points pertaining to liver function tests, including metrics like Total Bilirubin (TB), Direct Bilirubin (DB), Total Proteins (TP), Albumin (ALB), A/G ratio, as well as SGPT, SGOT, and Alkphos. The dataset consist of 583 patient records and this dataset includes records of 416 patients with liver issues and 167 patients without liver complications. These attributes represent basic blood tests utilized for gauging enzyme, protein, and bilirubin levels in the bloodstream, aiding in the identification of liver impairment. Proteins, essential for overall well-being, are large molecules, while enzymes act as crucial protein cells that facilitate vital chemical reactions within the body. Bilirubin assists in the breakdown and digestion of fats. The liver synthesizes crucial enzymes, namely ALT (SGPT), AST (SGOT) and ALP. ALT, AST, and ALP are specific liver enzyme tests employed to measure the levels of corresponding substances in the blood. Elevated ALT and AST levels may indicate potential liver damage, while heightened ALP levels might signal liver or bile duct harm.

Proposed Model

The emergence of machine learning (ML) technology in healthcare presents a promising alternative for the early diagnosis of liver diseases. ML algorithms can sift through vast amounts of medical data to identify patterns that might be missed by human practitioners. By leveraging historical patient data, these algorithms can accurately predict the likelihood of liver disease. Typically, machine learning-driven prediction of liver disease involves the use of structured datasets with medical attributes such as age, gender, bilirubin, and enzyme levels. The prediction workflow encompasses several critical stages: data collection, preprocessing, feature selection, model training, evaluation, and deployment. Among the preprocessing steps, addressing missing data, normalizing, and encoding variables is essential. Feature selection is also pivotal as it determines which attributes are most relevant for ensuring accurate predictions.

A variety of machine learning algorithms are utilized in this predictive context, each presenting distinct advantages and challenges. Random Forests enhance predictive accuracy by aggregating the results of multiple decision trees, thus operating as an ensemble method. Support Vector Machines (SVM) are particularly effective in dealing with high-dimensional data and can incorporate non-linear relationships through the use of kernel functions. Gradient Boosting algorithms, such as XGBoost, are noted for their performance in handling complex datasets. The selection of the most suitable algorithm is governed by the specific dataset and problem context, illustrating the flexibility and adaptability of machine learning methods in liver disease prediction.

Moreover, machine learning tools are adept at handling the prediction and classification of large datasets, with classification processes serving as a form of data mining aimed at uniting similar data or patterns into cohesive groups. This distinguishes machine learning from traditional programming paradigms, as machine learning focuses on pattern recognition and behavioral analysis within extensive datasets. Classification can also be complex due to the dynamic nature of frequently updated data, with various methods including supervised, unsupervised, and semi-supervised classification approaches existing to address these challenges.

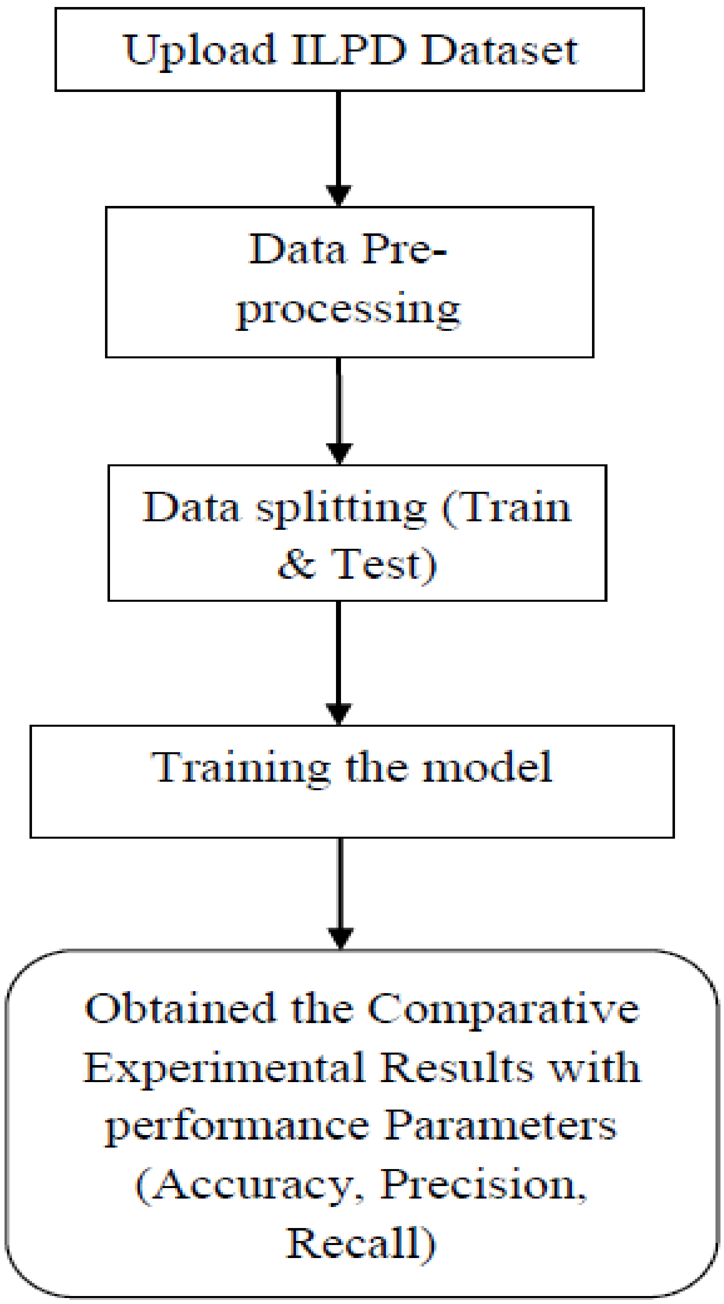

Figure 2: Proposed model for liver disease classification using machine learning classifier.

Table 2: This table presents comparative experimental results using different machine learning classifier.

Dataset name	Method	Accuracy
Indian Patients Liver Disease	**K-Nearest Neighbor**	**66**
	Random forest	**78**
	Extra Tree Classifier	**74**
	Support Vector Machine	**76**

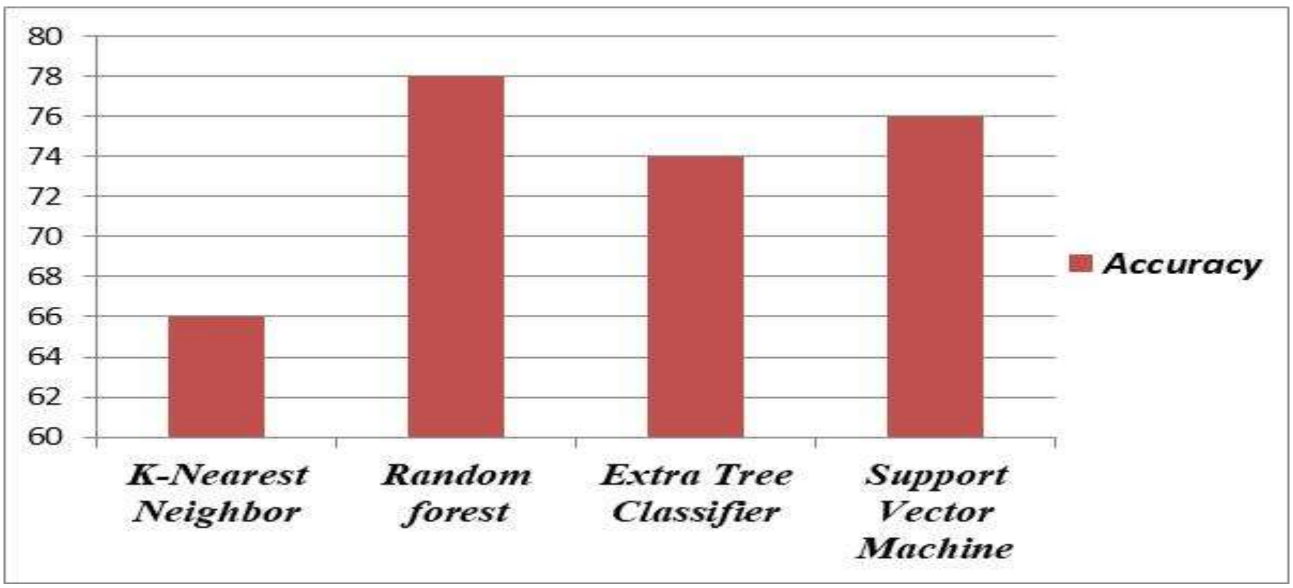

Figure 3: This picture shows that comparative experimental results using different machine learning classifier with performance parameters is accuracy.

Conclusion

Machine learning significantly differs from traditional programming languages by enabling pattern and behavioral analysis of large datasets. It categorizes data based on various attributes such as shape, size, and usage context. The classification of data presents challenges due to its dynamic nature, leading to the development of various approaches: supervised classification, unsupervised classification, and semi-supervised classification. This research focuses on the application of different classification models for diagnosing liver disease, employing supervised machine learning models that utilize the same dataset to assess results based on accuracy. Among the tested models, the random forest classifier achieved the highest accuracy, demonstrating the capability of machine learning techniques to derive insights based on historical data and specific dataset patterns.

References

[1] Harshita Rajpoot, Aaditi Raj, Sachin Bhoite, "Leveraging Machine Learning for Early Detection and Prevention of Liver Disease Mortality in India", Grenze International Journal of Engineering and Technology, 2025, pp. 1-8.

[2] Nidhi Arora, Shilpa Srivastava, "An enhanced predictive modelling framework for highly accurate non-alcoholic fatty liver disease forecasting", Indonesian Journal of Electrical Engineering and Computer Science, 2025, pp. 214-222.

[3] Nilam Solanki, "Machine Learning and Ensemble Learning Based Predictive Modeling for Cirrhosis Stage Classification", Procedia Computer Science, 2025, pp. 1452-1462.

[4] Gangfeng Zhu, Yipeng Song, Zenghong Lu, Qiang Yi, "Machine learning models for predicting metabolic dysfunction-associated steatotic liver disease prevalence using basic demographic and clinical characteristics", Journal of Translational Medicine, 2025, pp. 1-14.

[5] Yaran Chen, Xueyu Chen, "Multimodal Learning-based Prediction for Nonalcoholic Fatty Liver Disease", Machine Intelligence Research, 2025, pp. 871-887.

[6] Pankaj Kawadkar, Deepak Kumar Rathore, Devendra Singh Kushwaha, B Rebecca, G Priyanka Jeeva Karunya, PM Yohan, "Performance analysis of image compression techniques for MRI image using machine learning techniques", AIP Conference Proceedings, 2024, 10.1063/5.0239127.

[7] Ketan Gupta, Nasmin Jawani, Neda Afreen, "Liver Disease Prediction using Machine learning Classification Techniques", IEEE, 2022.

[8] 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 Volume 8, Issue 11, November 2019.

[9] Nazim Razali1, Aida Mustapha, Mohd Helmy Abd Wahab, SalamaA Mostafa1, Siti Khadijah Rostam, "A Data Mining Approach to Prediction of LiverDiseases", Journal of Physics, 2020.

[10] Pushpendra Kumar, and Ramjeevan Singh Thakur, "An Approach Using Fuzzy Sets and Boosting Techniques to Predict Liver Disease", Computers, Materials & Continua Tech Science PressDOI:10.32604/cmc.2021.016957.

[11] Deepak Kumar Rathore, and Sunil Phulre, "Machine Learning Based Efficient Prediction of Human Heart Disease by Identifying the Features", International Journal of Innovative Research in Technology and Management, (2022), pp. 21-27.

[12] Jingwei Wei,Hanyu Jiang,Dongsheng Gu,Meng Niu,Fangfang Fu, Yuqi Han, Jie Tian, “Radiomics in liver diseases: Current progress and futureopportunities”, Liver International, 2020.

[13] Mandakini Priyadarshani Behera, Archana Sarangi, Debahuti Mishra, Shubhendu Kumar Sarangi, “A Hybrid Machine Learning Algorithm for Heart and Liver Disease Prediction using Modified Particle Swarm Optimization”, International Conference on Machine Learning and Data Engineering, 2023.

[14] Se-Yeol Rhyou and Jae-Chern Yoo, “Cascaded Deep Learning Neural Network for Automated LiverSteatosis Diagnosis Using Ultrasound Images”, Journal of Sensors, 2021.

[15] Monica Lupsor-Platon , Teodora Serban, Alexandra Iulia Silion, George Razvan Tirpe, Alexandru Tirpe and Mira Florea, “Performance of Ultrasound Techniques and the Potential of Artificial Intelligence in the Evaluation of HepatocellularCarcinoma and Non-Alcoholic Fatty Liver Disease”, Journal of Cancer Treatment, MDPI.

[16] D Rathore, P Mannepalli, “A Review of Machine Learning Techniques and Applications for Health Care “, International Conference on Advances in Technology, Management & Education, 2021, IEEE proceeding, 978-1-7281-8586-6/21.

[17] D.Devikanniga, Arulmurugan Ramu and Ananda kumar Haldorai, “ Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crows Search Algorithm”, EAI Endorsed Transactions on Energy Web, 2020.

[18] Mazhar Javed Awan, “Review on Effective Disease Prediction through Data Mining Techniques”, International Journal on Electrical Engineering and Informatics · September 2021.

[19] Binish Khan, Piyush Kumar Shukla, “Strategic Analysis in Prediction of Liver Disease Using Different Classification Algorithms”, International Journal of Computer Sciences and Engineering · July 2019.

[20] D Rathore, Praveen Kumar Mannepalli, “Recent Trends in Machine Learning for Health Care Sector”, International Journal of Innovative Research in Technology and Management, Vol-5, Issue-2, 2021.

[21] Jaime Lynn Speiser, Constantine J. Karvellas, Bethany J. Wolf, Dongjun Chung, “Predicting daily outcomes in acetaminophen-induced acute liver failure patients with machine learning techniques”, Computer Methods and Programs in Biomedicine, 2019.

[22] Nazim Razali1, Aida Mustapha, Mohd Helmy Abd Wahab, SalamaA Mostafa1, Siti Khadijah Rostam, “A Data Mining Approach to Prediction of LiverDiseases”, Journal of Physics, 2020.