# Exploring Machine Learning Approaches in Liver Disease Diagnosis

**Rohini Temkar[1], Gresha Bhatia[2], Vidhi Chijwani[3], Ritesh Tahilramani[4], Nikita Narwani[5], Navin Idnani[6]**

[1,2,3,4,5,6] Department of Computer Engineering, VESIT, Mumbai, Maharashtra, India

[1] rohini.temkar@ves.ac.in , [2] gresha.bhatia@ves.ac.in

**Article Info**

***Article type:***
*Research*

***Article History:***

*Received: 2024-03-12*

*Revised: 2024-05-15*

*Accepted: 2024-06-25*

***Keywords:***
Decision tree, Support vector machine, Machine Learning(ML), XGBoost

**ABSTRACT**

Worldwide it has been observed that the count of patents with liver diseases has been consequently becoming widespread in the context of polluted environment, inhale of polluted gas, overuse of some drugs, packaged and contaminated foodstuffs and majorly drinking addictions., 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 an expert system for medicine located in a distant location will come in handy. The motive of this study is to evaluate the effectiveness and performance of various machine learning algorithms in predicting chronic liver disease, with the aim of reducing the high costs associated with its diagnosis. A range of machine learning approaches are presented in a comprehensive survey as a part of this work. These approaches are of prime importance in the diagnosis of liver disease. The notable machine learning algorithms used in this work demonstrate different degrees of sensitivity, precision and accuracy. The primary objective of this work is twofold: first, to give a complete survey of existing techniques on liver disease prediction; second, to carry out a comparative analysis of performance of the machine learning algorithms used in performance analysis. Through synthesizing insights from different algorithms, the goal is to enrich the knowledge of efficient methods for liver disease diagnosis and prediction in the medical field.

## 1. INTRODUCTION

As per World Health Organization records in 2023, liver disease accounts for 2.4% of all deaths in India. It is estimated that between 10 and 15 percent of people have liver disease, with rural areas having a higher frequency. One of the most common liver disease-related causes of death globally is cirrhosis. Every year, almost two million people die from liver illnesses, one million from cirrhosis complications, and another million from hepatocellular carcinoma and viral hepatitis. The liver is about a football size organ, right beneath the rib cage. An essential organ for food digestion and toxin elimination from the body is the liver. Genetic transmission is the cause of liver disease. Liver damage can also result from many liver-damaging conditions, such as infections, alcohol misuse, and obesity. Cirrhosis, a potentially fatal disease, can develop from liver damage over time and lead to liver failure. However, prompt treatment might help the liver heal. Early diagnosis may protect the liver from harm. The liver is a remarkable organ. Your liver can heal itself if the diagnosis is made after some scar tissue has already formed. Therefore, with the right disease management and treatment, liver disease damage can mostly be reversed. A liver transplantation procedure is required in some cases. Finding liver illness early on is an extremely difficult endeavour. This is also true even when the liver is injured. Fig. 1 depicts various stages of liver damage.

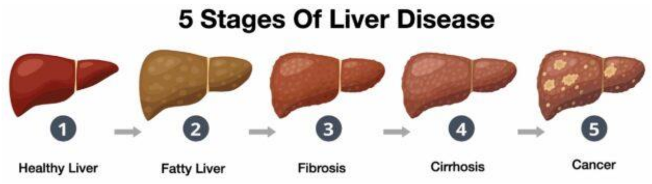


**Fig.1 stages of Liver Damage**
(source: https://www.reverseyourfattyliver.com)

Early prediction is crucial in providing appropriate treatment and saving patients' lives because medical expert systems can struggle to accurately diagnose diseases, leading to ineffective treatments and medications. Chronic liver diseases present a variety of symptoms, including digestive issues, jaundice, abdominal pain, dry mouth, dark-colored urine, loss of appetite, constipation, and internal bleeding. Dermatological symptoms may include yellowish skin, spider veins, acne, hives, rosacea, blisters, and redness in the feet. Additionally, neurological problems such as memory loss, numbness, and fainting may also occur. It's critical to take a number of preventive steps, including routine medical check-ups, vaccinations, cutting down alcohol usage, regular exercise, and weight management so that liver related issues can be prevented. Because they make liver disease detection and prognosis more accessible, the current medical liver disease diagnosis expert systems have benefited society. The accuracy and caliber of liver disease diagnosis and forecasting are being improved by the ongoing developments in artificial intelligence, especially the creation of diverse machine learning algorithms. As a result, early liver disease detection is essential for prompt treatment and recovery. It is still quite difficult to correctly identify liver disease in its early stages, though.

## 2. LITERATURE SURVEY

Table 1: Comparison amongst ML approaches in Liver diagnosis

| Sr. No | Reference | ML Algorithm used | Dataset used | Remarks (Accuracy/precision) |
|---|---|---|---|---|
| 1 | Amin, Ruhul, et al. [1] | KNN, logistic regression, SVM, multilayer perceptron | The data from University of California and Irvine (UCI) repository related Indian liver patients | The system's performance in predicting liver disorders was 88.10% accuracy, 85.33% precision, 92.30% recall, 88.68% F1 score, and 88.20% AUC score. |
| 2 | Tokala, Srilatha, et al. [2] | SVM,KNN,ANN and NB | UCI ILPD Dataset | Despite increasing health awareness, liver and heart diseases are rising. This project proposes a highly accurate AI system (100% for SVM with smaller datasets) to predict liver disease risk exceeding 90% accuracy. and Naive Bayes with lowest accuracy of 55.6%. |
| 3 | Sreenivasa Rao Veeranki and Manish Varshney [3] | Random Forest, Knn, SVM, MLP models | Indian liver Patient dataset | The SVM model achieved an accuracy of 72%, while Random Forest, KNN, and MLP models achieved accuracies of 69%, 55%, and 64%, respectively. |
| 4 | Ketan Gupta et al. [4] | SVM, LR, RF, k-NN | Chronic liver disease dataset | SVM achieved highest accuracy of 91.7% |

| | | | | |
|---|---|---|---|---|
| 5 | Taher M. Ghazal et al. [5] | SVM, RF, k-NN, NB | A dataset from UCI MLRepository | SVM achieved highest accuracy of 93.75% |
| 6 | Behzad Hatami et al.[6] | KNN, SVM,and neural network | A Dataset from SB University of Medical Sciences' Research Institute | The top performing method was determined to be KNN, which had an accuracy rate of 98% on average. RF performed the best after that. |
| 7 | Chauhan, Bhavana Singh, et al. [7] | LR,RF,SMO,J.48,NB,IBk | Own clinical dataset | SVM achieved an accuracy of 90%. |
| 8 | Shubashini Rathina Velu,Vinayakumar Ravi & Kayalvily Tabianan [8] | Decision Tree and Naive Bayes | A Dataset of Liver Disease Patients from Kaggle | By using data mining and a classification model, the DT model beats the other models and , achieving 98.40% accuracy . |
| 9 | K Jeyalakshmi and R Rangaraj [9] | Modified Convolutional neural network(CNN_LDPS),MLPNN, | UCI ML Repository | Proposed CNN-LDPS has improved accuracy than MLPNN |
| 10 | Mostafa, F., et al [10] | RF, SVM, ANN | UCI ML Repository & own clinical dataset | RF achieved the highest accuracy (98.14%). |
| 11 | Ishtiaqe Hanif and Mohammad Monirujjaman Khan [11] | RF,SVM,DT | UCI ML Repository | RF and SVM both algorithms achieved state-of-the-art accuracy 97% and DT with lowest accuracy of 75%. |
| 12 | Golmei Shaheamlung and Harshpreet Kaur [12] | CNN, SVM, LR, Decision Tree, k-NN | Chronic liver disease dataset | CNN achieved highest accuracy of 92.75% |

## 3. MACHINE LEARNING IN HEPATOLOGY

Hepatology is a branch of medicine dedicated to the study, diagnosis, and treatment of conditions affecting the liver, gallbladder, pancreas, and biliary system. Machine learning. The following algorithms are used for liver disease diagnosis.

2.1. Decision Tree

The relentless quest for early and accurate liver disease diagnosis has led to the exploration of various advanced tools. Among these, machine learning algorithms have emerged as formidable contenders, exhibiting remarkable accuracy and offering promising avenues for improved healthcare outcomes. Let's unveil the unique capabilities of four key algorithms that shine brighter than their peers in this challenging battle: Put yourself into an exacting role of a detective who is examining the patient data and formulating a set of questions. One of the advantages of this method known as Decision Tree is that not only can it offer a lot of accuracy in diagnosis, but also its reasoning is very understandable unlike other methods that are often difficult to interpret. Although this technique could tap bias and thus be not so efficient its efficiency could be decreased due to the fact that it relies on only one decision path thus the smaller the dataset the more likely it is to decrease the efficiency of the machine.

2.2. Random Forest

One can imagine a tribunal of decision trees, each trained on a different subset of the data, that would vote to reach a collective decision. The ensemble strategy types called "Random Forest" use multiple individual results which the members of these systems calculate, interchanging high accuracy with their robustness. Just an assortment of laws in a complex or unpredictable legal system, the granular integration among trees may not be easily understood by its internal functioning. Random Forest Classifier, utilizing SMOTE-ENN balancing, is effective for liver disease prediction by handling data imbalance and leveraging decision tree ensembles for accurate diagnosis.

2.3. Support Vector Machine (SVM)

Imagine a data strategist drawing a decisive line at a battlefield, dividing health from a disease stage at the data. SVMs have the advantageous property of capturing even more complicated patterns, that is why they are the most powerful methods to deal with intricate diagnostic tasks. Notwithstanding, the optimization of the algorithms' parameters is vital to achieve in the best exercise of this tool.

2.4. XGBoost

One well-liked and effective technique for a variety of machine learning problems, including classification, is called XGBoost, which is an implementation of gradient boosting. Based on a set of features, the XGBoost classifier is used in the provided code to predict the condition of liver cirrhosis. The model's accuracy is evaluated using the accuracy score metric after it has been trained on the training dataset and has made predictions on the test dataset. Liver cirrhosis is accurately predicted by the XGBoost classifier. Given the importance of reliable and precise forecasts for informed clinical decision-making, XGBoost's excellent accuracy indicates that it is especially helpful in medical applications such as liver disease prediction.

## 4. METHODOLOGY

The suggested approach uses a structured pipeline for machine learning. Before testing and training, obtained metrics like ascites, hepatomegaly, spiders, and biochemical markers are processed as part of the pipeline's first step, data preparation, to improve quality. Normalization, feature extraction, and handling missing values are some of the pre-processing procedures used to guarantee the best possible quality of data.

### *4.1. Workflow of the system*

The workflow of the system is depicted in Fig. 2 and the steps are described as follows:

1. **Data Collection:** Gathering dataset with twenty as main features; from that, utilized thirteen features only that have more importance.
2. **Data Pre-processing and Cleaning:** This block involves tasks like dropping missing values, noise reduction, and data augmentation to prepare the dataset for training. Data cleaning ensures high-quality input for the model.
3. **Model Training:** Train different models such as Logistics regression, random forest classifiers, and decision tree using pre-processed data The training process involves feature extraction and model optimization.
4. **Model Evaluation:** Using validation datasets and metrics like accuracy, precision, recall, and F1-score, evaluate the model's performance. Adjust the model in light of the evaluation's findings.
5. **Deployment:** Deploying the model to production environments using flask.
6. **Integration with web application:** Ensuring seamless integration of the Plant Disease Prediction Model with the Mobile and Web Application. Developing APIs or endpoints for communication between the application and the model.
7. **Input image from user:** Implement functionality in the application to allow users to submit their data for analysis. Process user input for prediction.
8. **Final output to the user:** Present the analysis results to users in an understandable format. Include information about the detected disease, severity, and recommended actions. Present the analysis results to users in an understandable format. Include information about the detected disease.

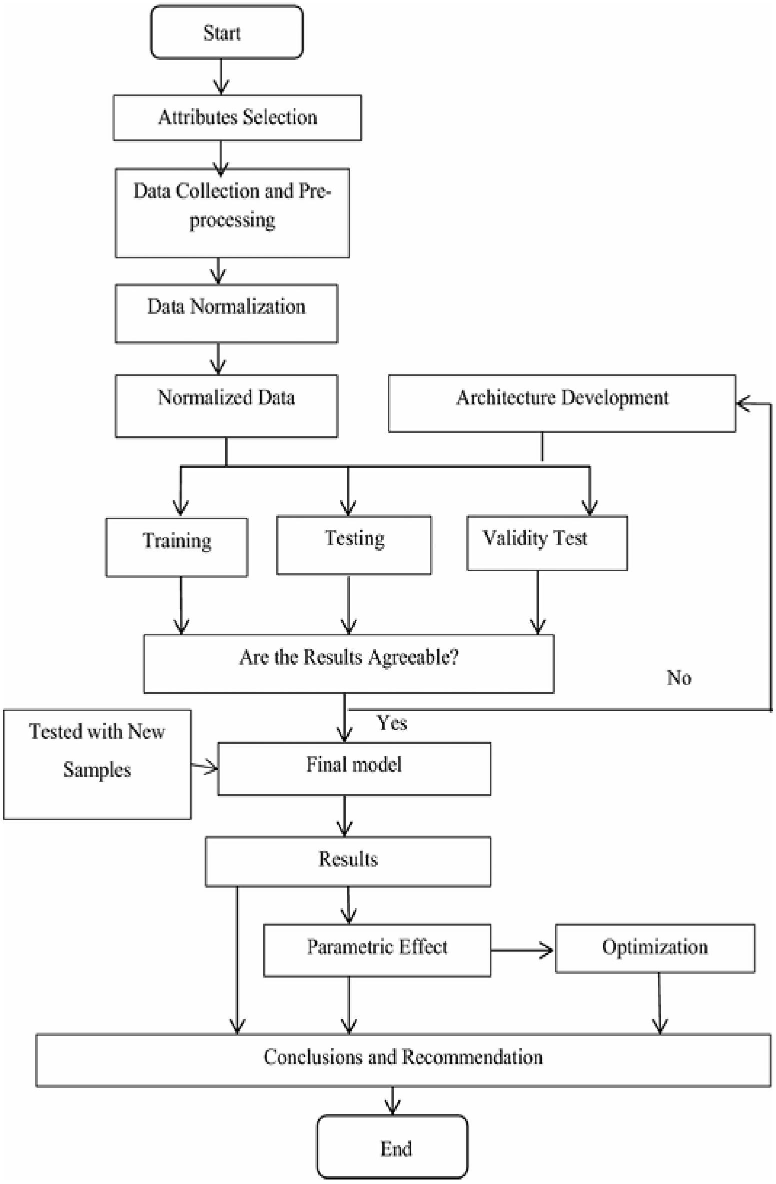


**Fig.2 Workflow of the system**

### *4.2. Input parameters*

The input parameters considered for evaluation purpose are as follows:

1. **Ascites:** An accumulation of fluid in the abdomen, frequently accompanied by cirrhosis or other advanced liver disease. Low blood protein levels and elevated hepatic blood vessel pressure (portal hypertension) are the causes of it.
2. **Hepatomegaly:** Liver enlargement, frequently observed in cirrhosis and other liver disorders. It can be brought on by inflammation, fatty liver alterations, or liver tissue scarring (fibrosis).
3. **Spiders:** People with liver cirrhosis frequently have spider angiomas or spider nevi, which are tiny, dilated blood veins close to the skin's surface. They are brought on by modifications in hormone levels and blood flow brought on by liver disease.
4. **Edema**: Fluid retention-related swelling, usually in the abdomen or legs. Edema may develop in liver cirrhosis as a result of the liver's reduced ability to produce protein, which lowers blood albumin levels and raises blood vessel pressure.
5. **Bilirubin:** A yellowish pigment generated

during the lysis of red blood cells. Since the liver is in charge of processing and excreting bilirubin, elevated blood bilirubin levels (hyperbilirubinemia) may be a sign of liver malfunction.

6. **Cholesterol:** A blood-borne fat type that is linked to a higher risk of cardiovascular disease at high levels. Abnormal lipid profiles can result from liver cirrhosis's impact on cholesterol metabolism.
7. **Albumin:** A liver-produced protein that carries a variety of materials, including hormones and prescription drugs, and aids in preserving blood's fluid balance. Because liver function is compromised in liver cirrhosis, low albumin levels, or hypoalbuminemia, are frequently observed.
8. **Copper:** A vital trace mineral that is used in several different metabolic processes. A malfunction in copper metabolism in liver cirrhosis can result in an abnormal build-up of copper in the liver and other organs (Wilson's disease).
9. **Alk_Phosph:** The enzyme alkaline phosphatase is present in the liver, bones, and bile ducts, among other tissues. Increased blood levels of alkaline phosphatase could be a sign of bone or liver illness, particularly cirrhosis of the liver.
10. **Serum Glutamic Oxaloacetic Transaminase, or SGOT:** This enzyme, which is present in the liver and other tissues, is also referred to as AST (aspartate aminotransferase). Increased blood levels of SGOT may be a sign of inflammation or injury to the liver, both of which can happen in liver cirrhosis.
11. **Tryglicerides:** High triglyceride levels may indicate fatty liver disease, which is frequently caused by poor dietary habits. Fatty liver disease typically does not cause symptoms, but it can result in severe liver damage and cirrhosis.
12. **Platelets:** Blood components that aid in coagulation. Due to decreased production by the diseased liver or greater sequestration in the spleen, platelet counts may drop in liver cirrhosis, increasing the risk of bleeding.
13. **Prothrombin:** The duration of blood clotting is measured by prothrombin time (PT). The liver's capacity to manufacture clotting factors may be compromised by liver cirrhosis, which may result in longer parenchyma therapy and a higher risk of bleeding.

## 4. PERFORMANCE ANALYSIS

1. **Accuracy:** It is the percentage of accurately anticipated cases (both positive and negative) out of all examples analyzed. Accuracy is calculated as: Accuracy = (True Positive + True Negative) / (True Positive + True Negative + False Positive + False Negative)
   Explanation: A high accuracy score means that the model minimizes misclassifications by properly predicting liver cirrhosis cases as positive and non-cirrhosis cases as negative.
2. **Recall:** Recall is a machine learning parameter that assesses how well a model can detect positive instances in a dataset.. It is also known as sensitivity. Recall is calculated as: Recall = True Positives / (True Positives + False Negatives). In simpler terms, recall indicates the proportion of actual positive instances (cases of liver cirrhosis) that the model correctly identifies. A high recall value suggests that the model is effective at capturing most of the positive cases, minimizing the number of cases missed (false negatives).
3. **Precision:** Precision is a measure of how accurate the model's positive predictions are. It is calculated as: Precision = True Positives / (True Positives + False Positives). The precision of a model is measured by the percentage of true positive predictions (i.e., accurately detected cases of liver cirrhosis) among all positive predictions. A high precision score indicates that the model makes accurate positive predictions while reducing false positive predictions. Precision is especially beneficial when the consequences of false positives are severe, because it ensures that the model's positive predictions are precise and trustworthy.

Table 2: comparisons amongst different machine learning techniques for liver disease detection based on performance

| Methods | % Accuracy | % Sensitivity | % Precision |
|---|---|---|---|
| Logistic Regression | 75 | 78 | 78 |
| Decision Tree | 95 | 95 | 95 |
| SVM | 87 | 86 | 85 |
| XGBoost | 98 | 97 | 96 |

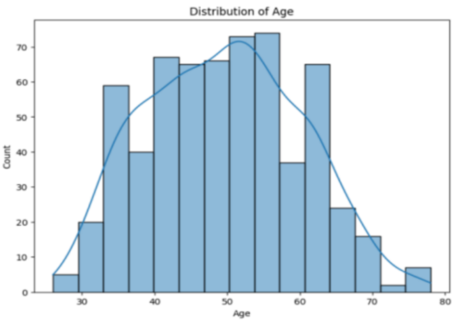


**Fig. 3 Distribution of Age**

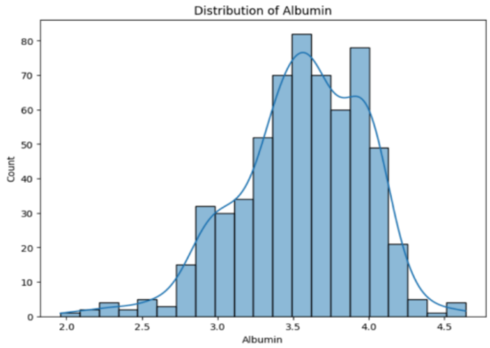


**Fig. 4 Distribution of Albumin**

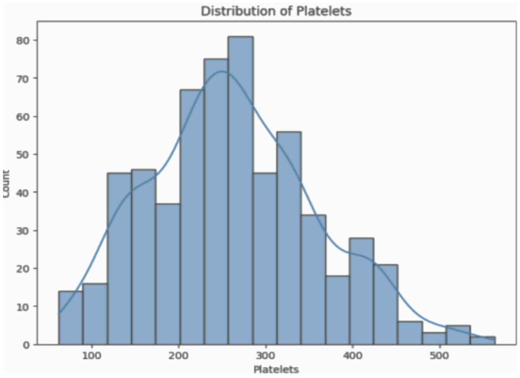


**Fig. 5 Distribution of Platelets**

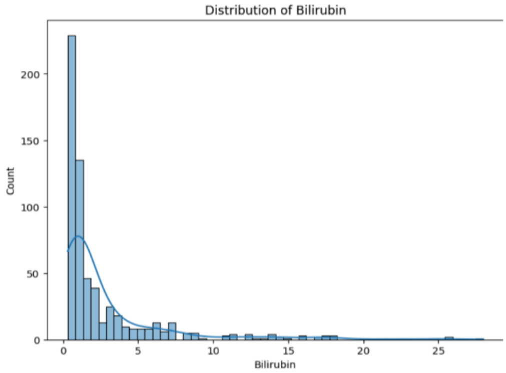


**Fig. 6 Distribution of Bilirubin**

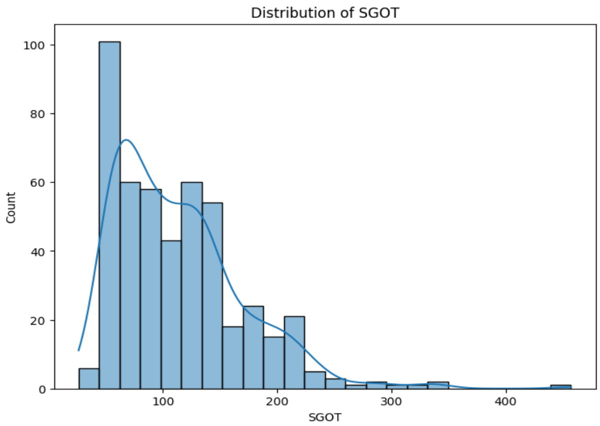


**Fig. 7 Distribution of SGOT**

## 5 . CONCLUSION

The exciting potential of machine learning in the diagnosis of liver illness is highlighted in this research. Each composition has its strengths and in individual studies the "best performer" changes, for example, in accordance to dataset properties, selected features and the particular task in question. This emphasizes the fact that one should be very selective when picking and having a critical evaluation of algorithms for any particular case. Through the implementation of Random Forest, XGBoost, Support Vector Machine, and Logistic Regression models, the study demonstrated the efficacy of XGBoost in achieving an accuracy of 98%, outperforming other algorithms. With the use of sophisticated machine learning methods and extensive analysis of patient data, this system presents a viable answer to the ongoing problem of early liver cirrhosis identification. By forecasting the likelihood of cirrhosis and allowing for early intervention, this technique has the potential to minimize the healthcare costs and reduce the burden on healthcare systems.

## REFERENCES


[1] Amin, Ruhul, et al. "Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms." Informatics in Medicine Unlocked 36 (2023): 101155.

[2] Tokala, Srilatha, et al. "Liver disease prediction and classification using machine learning techniques." International Journal of Advanced Computer Science and Applications 14.2 (2023).

[3] Veeranki, Sreenivasa Rao, and Manish Varshney. "Intelligent techniques and comparative performance analysis of liver disease prediction." International Journal of Mechanical Engineering 7.1 (2022): 489-503.

[4] K. Gupta, N. Jiwani, N. Afreen and D. D, "Liver Disease Prediction using Machine Learning Classification Techniques," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 2022, pp. 221-226, doi: 10.1109/CSNT54456.2022.9787574.

[5] Ghazal, Taher M., et al. "Intelligent model to predict early liver disease using machine learning technique." 2022 International Conference on Business Analytics for Technology and Security (ICBATS). IEEE, 2022.

[6] Hatami, Behzad, et al. "Machine learning-based system for prediction of ascites grades in patients with liver cirrhosis using laboratory and clinical data: design and implementation study." Clinical Chemistry and Laboratory Medicine (CCLM) 60.12 (2022): 1946-1954.

[7] Chauhan, Bhavana Singh, et al. "Effect of etanercept on Plasmodium yoelii MDR-induced liver lipid infiltration." Future Pharmacology 2.4 (2022): 499-510.

[8] Velu, Shubashini Rathina, Vinayakumar Ravi, and Kayalvily Tabianan. "Data mining in predicting liver patients using a classification model." Health and Technology 12.6 (2022): 1211-1235.

[9] Jeyalakshmi, K., and R. Rangaraj. "Accurate liver disease prediction system using convolutional neural network." Indian Journal of Science and Technology 14.17 (2021): 1406-1421.

[10] Mostafa, F., et al. "Statistical Machine Learning Approaches to Liver Disease Prediction. Livers 2021, 1, 294–312." (2021).

[11] Hanif, Ishtiaqe, and Mohammad Monirujjaman Khan. "Liver cirrhosis prediction using machine learning approaches." 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2022.

[12] G. Shaheamlung, H. Kaur and J. Singla, "A Comprehensive Review of Medical Expert Systems for Diagnosis of Chronic Liver Diseases," 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 2019, pp. 731-735, doi: 10.1109/ICCIKE47802.2019.9004438.

[13] K. Shalini and K. Karthikeyan. "Liver disease prediction using machine learning algorithms." International Journal of Scientific & Technology Research 8.12 (2019): 245-250.

[14] Snehal B. Patil and Ramesh R. Manza. "Liver disease prediction using machine learning techniques." International Journal of Engineering Research and General Science 7.2 (2019): 156-161.

[15] Avneet Kaur and Dr. Gaganpreet Kaur. "A comparative study of machine learning algorithms for liver disease prediction." International Journal of Advanced Research in Computer Science and Software Engineering 8.8 (2018): 72-78.