International Journal of Basic and Applied Sciences, 14 (4) (2025) 522-532

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/vg1c4b90
Research paper

Harmonizing Machine Learning Algorithms for Enhanced Liver Disease Prediction: A Comparative Study of Accuracy and Efficacy

Laxminath Tripathy [1] *, Dharmveer Kumar Yadav [2], Brijesh Kumar Singh [3]

[1] Faculty of Engineering & Technology, ITER, Siksha 'O' Anusandhan University. Odisha, India
[2] Katihar Engineering College, Katihar, Science, Technology and Technical Education Department, Govt. of Bihar, India
[3] Govt. Engineering College, Jehanabad, Science, Technology and Technical Education Department, Govt. of Bihar, India
**Corresponding author E-mail: laxminathtripathy@soa.ac.in*

Received: April 26, 2025, Accepted: June 27, 2025, Published: August 19, 2025

Abstract

In this review paper, various machine learning techniques applied to predict liver disease have been studied. Non-Alcoholic Fatty Liver Disease (NAFLD) is a leading cause of global mortality, which affects millions of people worldwide. The diagnosis of NAFLD can be both costly and complex. Therefore, this research aims to reduce these challenges by assessing the capability of various machine learning models to prevent NAFLD, thereby reducing the overall cost of diagnosis. We have used 3 classification algorithms-Logistic Regression, Support Vector Machine, along the Random Forest technique for enhanced prediction accuracy. Our study evaluates these models based on metrics like accuracy, precision, recall, and F1 score. Also, we are inclined towards developing a hybrid model to further improve efficiency. Based on our study of various other papers, it was found that integrating-ing multiple Machine Learning (ML) techniques helps to significantly enhance the performance, offering an efficient approach towards the diagnosis.

Keywords: *Liver Disease; Machine Learning; Classification; Feature Selection; Ensemble Techniques; Support Vector Machines; Random ForEst; Logistic Regression.*

1. Introduction

In the current scenario, where every person meets various diseases, more than a million people are diagnosed with liver disease each year. Liver disease problems like liver cirrhosis, hepatitis (A, B, C), and liver cancer are some of them. On a global level, there are 1.32 million more deaths from liver cirrhosis in 2017 than in 1990, of which 66.7% were men and 33.3% were women. The largest gland in our body is the liver (from the Greek word hepar, meaning liver), which is located on the right side of the upper abdominal cavity. The pancreas has both exocrine and endocrine functions. The liver performs numerous metabolic functions that are essential for maintaining homeostasis, supporting nutrition, and enabling immune responses. Its main functions are:

1) Bile secreting with nutrient storage.
2) It, except for serum protein, also produced lipids.
3) Detoxifies blood with various endogenous and exogenous substances (toxins, drugs, alcohols, etc) in the circulation.
4) During fetal life, it forms hemopoietic cells of all types. The liver originates from a diverticulum (liver bud) of the distal foregut.

The liver bud elongates cranially (towards the head) and extends to the right, giving rise to a lateral outgrowth, the accessory bud pars cystic (bud on the right side of the common channel, which forms— Mainly the cystic duct and gallbladder). Pars livera, the principal bud, develops into the septum transversum. It splits distal to the liver, bifurcating into the right and left liver ducts and liver parenchyma. Before the liver bud invades, the vitelline veins and umbilical veins open into the septum transversum. These vessels branch and serve as the source of true sinusoids that infiltrate into the liver parenchyma and divide it into liver cords. Bile canaliculi and ductules link the liver parenchyma, and the extraliver bile ducts are developed in a subsequent step. Bilirubin metabolism refers to the process by which the liver handles waste products like bilirubin, breaking them down and excreting them into the gut. Most of the bilirubin (about 80–85%) is produced from the recycling of hemoglobin in senescent (aged) red blood cells. The remaining 15–20% of bilirubin is derived from heme breakdown through alternative pathways, including the cytochrome P450 enzyme system and the hemolysis of young red blood cells within the bone marrow. The initial form produced, unconjugated bilirubin, is a substance that is difficult for the body to process. To get around this, unconjugated bilirubin tags along in the bloodstream, as it is carried on a substance known as albumin, a transport protein in blood plasma. This albumin-bilirubin complex carries the waste product back to the liver, which is the main organ of bilirubin metabolism. Upon reaching the liver, the albumin-bilirubin complex dissolves and produces unconjugated bilirubin that penetrates the liver cells. This is when it goes through a bunch of vital changes. Unconjugated bilirubin first attaches to proteins within the cytoplasm of the cell. It then undergoes conjugation in which bilirubin conjugates with glucuronic acid, which renders it water-soluble. The conversion of this fat-soluble, poorly

transportable unconjugated bilirubin to water-soluble, easily excretable conjugated bilirubin (bilirubin diglucuronide) is necessary for the safe and efficient excretion of bilirubin from the body. This is The Last Step in Bilirubin Metabolism. Of course, the excretion. Conjugated bilirubin is secreted from the liver into the bile, which helps in digestion. The carcinogenic nature of the pathological destruction of erythrocytes that have undergone hemolysis due to some reasons causes two main substances, including bilirubin and biliary pigments, that are removed from the body after the liver excretes them into bile, and bile then passes through the bile ducts to the small intestine, and ultimately, conjugated bilirubin is expelled in feces. Gut bacteria also help metabolize bilirubin in the intestine.

There are existing predictive models for liver disease detection that have certain limitations, such as suboptimal accuracy, lack of generalizability, or reliance on a single algorithm that needs to be addressed. In recent years, we have witnessed rapid advancements in machine learning techniques and their successful application in various domains, including healthcare. Early and accurate prediction of liver disease is a major concern, and if done efficiently, can greatly improve patient outcomes by enabling timely interventions, preventive measures, and appropriate treatment plans. Early prediction of liver diseases can potentially reduce the financial burden on healthcare systems by preventing costly complications and hospitalizations. Overall, the motivation behind this work stems from the potential to improve patient outcomes, address the limitations of existing predictive models, enhance clinical decision-making, optimize healthcare resource utilization, and advance scientific understanding in the critical field of liver disease management—ultimately enabling earlier diagnosis and better care for patients. The shortcomings of traditional diagnostic approaches (e.g., liver biopsies and imaging) include their invasiveness, cost, and limited sensitivity in early-stage detection. These limitations can be overcome with the advantages offered by ML-based techniques.

Liver diseases have always had a huge impact on the global health index. This research aims to use the capabilities and power of machine learning algorithms to accurately and effectively predict liver disease conditions. It involves a comprehensive evaluation and comparison of various machine learning techniques like logistic regression, decision trees, random forests, support vector machines, SMOTE for dataset balancing, ensemble techniques, and hybrid architectures for enhanced efficiency. A major focus is on enhancing the overall prediction accuracy by addressing challenges such as data quality issues, class imbalance, model interpretability, and overfitting. Key steps include preprocessing patient data from sources like medical records and liver function tests, extracting the most predictive features, and rigorously validating the models. Beyond just accuracy, the research emphasizes evaluating models on other clinically relevant techniques like precision, recall, F1-score, and ROC-AUC. Ultimately, the objective is to create a superior predictive model that can provide timely and reliable risk estimates, thereby enabling more effective clinical interventions and management strategies for liver diseases. The various factors affecting the liver and the different stages of liver disease are illustrated in Fig. 1 and Fig. 2, respectively.

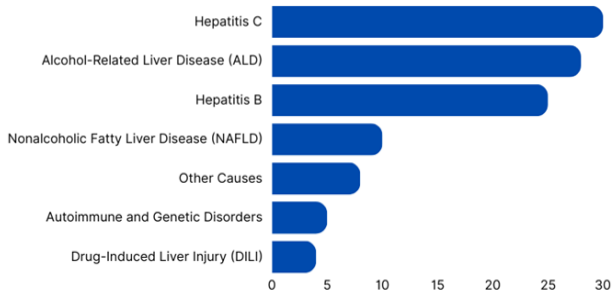

Fig. 1: Factors Affecting Liver.

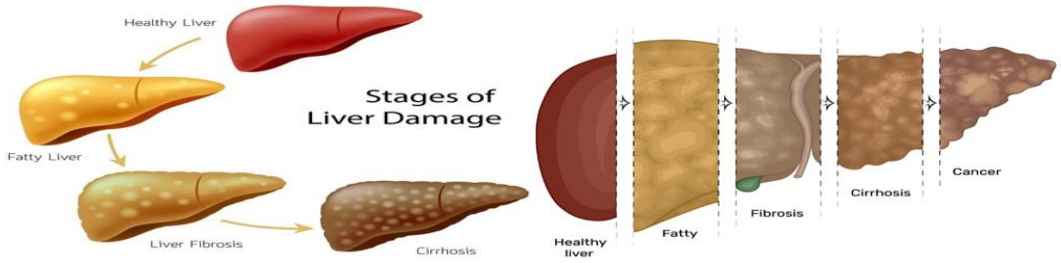

Fig. 2: Stages of Liver Disease.

This paper begins with Section 1: Introduction, which outlines the motivation for undertaking this research, emphasizing the need to address specific gaps and requirements identified in previous studies. It also states the main objectives of the study and provides an overview of the original contributions made.

Section 2 presents a literature review related to the topic, discussing key findings, methodologies used, and the gaps this study aims to fill. In Section 3, a general overview of the data sources and machine learning techniques employed is provided. Section 4 details the proposed system or model, explaining the methods and techniques used in its development. Section 5 covers experimentation and model evaluation, presenting the results through tables, graphs, and other visual aids. Finally, Section 6: Conclusion and Future Scope summarizes the findings, discusses their implications for the research problem, and suggests potential extensions or improvements for future work.

2. Literature survey

This literature review introduces machine learning techniques as a promising method for predicting and classifying liver diseases. In [1], the authors demonstrated that data mining combined with the Linear Discriminant algorithm can diagnose liver diseases with an accuracy of 95.8%, representing a significant advancement in medical diagnosis and treatment. ML algorithms for liver disease prediction using the Indian Liver Patient Dataset have been considered in [2], and this research mainly focuses on Naive Bayes over KNN. Utilizing six variables and highlighting Logistic Regression's high sensitivity, it also emphasizes liver disease prevalence and its impact among adults aged 40-60, particularly in India, where there are high mortality rates associated with these conditions.

Studies in [3] and [5] evaluate various machine learning techniques for predicting chronic liver disease using clinical data. According to [3], six algorithms were compared, with Logistic Regression achieving the highest accuracy of 75%. The study in [5] reports an accuracy of 97.12% using a Decision Tree model and further proposes an Artificial Neural Network achieving 99.9% accuracy. In [13], seven

algorithms were evaluated, with Random Forest identified as the most accurate at 83.7%, emphasizing the importance of early intervention in liver disease management.

In [4], the authors have discussed some factors leading to the escalation of liver diseases, like alcoholism and poor eating habits. This emphasizes the need for reliable detection methods to reduce healthcare costs. They have also highlighted complications related to liver diseases like NAFLD, which means Non-Alcoholic Fatty Liver Disease (it is a situation where excess fat accumulates in the liver especially related to people who consume alcohol in absolute amounts), and have proposed further study in this area. Therefore, it can be said that machine learning has succeeded in the accurate identification of liver diseases for better patient management.

In [6] and [12], the authors have focused on the various use cases of AI and ML in the sector of hepatology. In [6], the authors have focused on predicting the risk of liver disease using ML algorithms like LR and SVM, and implemented a system for use by the patient based on their blood test data. In contrast, Paper [12] provides a much broader aspect of the techniques of AI in the medical sector, i.e., hepatology, including various processes like Machine Learning and Deep learning techniques. While in [6] authors emphasize a specific implementation, [12] focuses on the overall impact of AI in liver disease prediction and management, thereby highlighting the futuristic directions and challenges.

In [7], the study aims particularly to evaluate classification models for predicting fatty liver disease (FLD) based on initial screenings at New Taipei City Hospital. It includes random forest (RF), Naïve Bayes (NB), artificial neural networks (ANN), and logistic regression (LR). RF has shown the highest performance with an AUROC of 0.925 and an accuracy of 87.48%, performing better than NB, ANN, and LR. The application of RF in clinics may help classify FLD patients for the early detection of liver cirrhosis.

In [8], the study proposes the use of an integrated feature extraction approach using ML for the classification of liver diseases with an accuracy of 88.10% thereby outperforming other existing methods. In [9] and [11], the authors have shown and depicted the marvelous effects of machine learning in Hepatology by showcasing its power to unveil and identify the hidden patterns for predictive modeling in liver diseases. In [9], the authors focus on the growing use of AI for diagnosis and treatment of liver disease, including ML algorithms, Deep learning algorithms, and the usage of NLP. [11] Specifically focuses on the usage of ML methods for the identification of significant classifiers using body composition and anthropometric variables. It conducts a study to assess various ML algorithms for the prediction of NAFLD, with RF having the most accuracy for fatty liver, steatosis stages, and fibrosis stages

Papers [10] and [13] explore the various applications of ML in the domain of liver disease prediction. Authors in [10] focus on using machine learning classifiers like SVM (Support Vector Machine), Boosted C5.0, and NB (Naive Bayes) in the prediction of liver diseases. The authors evaluate the classifier performance using metrics like accuracy, sensitivity, and specificity. In [13], the authors aim to predict and diagnose chronic liver disease by evaluating various machine learning algorithms, including logistic regression, random forest, XGBoost, and others. It is found that the random forest algorithm has achieved the most accurate results with superior performance in precision, recall, F1 score, and AUC metrics.

In [14], the purpose is to build Machine Learning models for early diagnosis of FLD, using SVM, Logistic Regression, RF, Naive Bayes, and MLP with Mutual Information feature selection. For dealing with imbalanced datasets, the Synthetic Minority Oversampling Technique (SMOTE) has been used. The results show that SVM performed the best among classifiers, attaining 99% accuracy in addition to RF. In summary, the research provides a development and comparative analysis of five models with a focus on SVM and RF, having higher accuracies in diagnosing FLD.

In this research paper [15], the authors explored how machine learning can be used to predict non-alcoholic fatty liver disease (NAFLD) in a group of 15,315 Chinese individuals. We developed seven different predictive models and assessed their performance using various measures. The XGBoost model stood out for its high accuracy (0.795), area under the receiver operating characteristic curve (AUROC) of 0.873, and area under the precision-recall curve (AUPRC) of 0.810. We also discovered that body mass index was the most important factor in predicting NAFLD. These results indicate that machine learning, especially the XGBoost model, shows promise in enhancing NAFLD screening and could lead to better early detection and treatment of the disease.

Both [16] and [17] discuss using machine learning algorithms to diagnose diseases. For example, [16] uses a multi-class dataset of 7 risk factors that span a decade for many people with an accuracy rate of 76% in its use of the J48 algorithm. Conversely, in the case of [17], the focus is on liver disease diagnosis due to its costliness as well as time-consuming aspect, and puts forward an intelligent model that has an accuracy of 0.884 and a miss rate of 0.116. Both studies emphasize the importance of early diagnosis for patients with various ailments, demonstrating how these approaches can make medical decisions more cost-effective and efficient. Nevertheless, the clustering in [16] considers multiple risk factors and classification, while that by [17] narrows down to liver disease detection only, thereby affirming that there are various ways through which machine learning techniques can be applied within the medical domain settings

Papers [18] and [20] address the challenge imposed for early detection of liver disease due to limited symptoms. Authors in [18] focus on the application of Medical Data Mining (MDM) using a Support Vector Machine (SVM) for the prediction of liver disease on the basis of historical data and patterns. Authors in [20] propose an architecture specifically related to ensemble learning and enhanced preprocessing for liver disease prediction. It was seen that high accuracy was achieved with ensemble learning algorithms such as Extra Tree Classifier and Random Forest. While both papers aim at improving the detection of liver disease in the early stages, Paper [20] provides a more efficient and detailed methodology and achieves higher accuracy with its proposed model.

In [19] the authors discuss how the medical imaging field has been transformed through the application of artificial intelligence (AI), especially algorithms like deep learning algorithms, specifically, in liver diagnostics. It boosts precision and speed in detecting liver complications using a variety of imaging methods by automating complex image analysis. The burden on physicians is reduced while diagnostic accuracy improves significantly. In this domain, such as traditional machine learning as well as deep learning, CNN is central. Visual diagnostics are augmented with AI's quantification, ensuring reproducible results. Meanwhile, challenges regarding deploying deep learning techniques into clinical practice, as well as future possibilities, were also addressed.

The authors in [21] emphasize the potential of ML (ML) algorithms, trained on key liver function variables like AST, ALT, GGT, BIL, and ALP, which showed high accuracy in diagnosing chronic liver disease. Factors such as ALT/AST ratio, GGT, ALP concentrations, and injury proximity to bile ducts were crucial in identifying disease causes and severity. SVM and RF algorithms outperformed ANN, with SMOTE enhancing minority group classification. Future work includes LIME for model interpretability and multinomial classification for different liver disease types. ML methods, while supporting medical experts, aren't replacements for clinical decisions but can address healthcare challenges, aid early detection, and reduce liver disease burden, potentially lowering mortality and invasive procedures.

The authors in [22] use data mining for the prediction of liver disease by analyzing the patient dataset. By normalization, feature selection technique with PSO, and application of classification algorithms, it is found that the J48 algorithm achieved the highest accuracy of 95.04%.

The study in [23] evaluates machine learning (ML) models for liver disease prediction using a dataset provided in the Indian Liver Patients' Records. While previous models achieved accuracies from 69.30% to 75.54%, this study focuses on ensemble learning, particularly the Voting method, which achieved 80.1% accuracy after Synthetic Minority Over-sampling Technique (SMOTE) and cross-validation. The

Voting method shows promise for accurate liver disease prediction. However, the study acknowledges dataset limitations and the need for more comprehensive medical data for further analysis. Future work aims to refine prediction methodologies using Region of Practical Equivalence (ROPE) analysis and deep learning techniques.

This article [25] aims to enhance liver disease diagnosis by examining patient parameters and genome expression. It evaluates computational algorithms for this purpose and proposes methods to improve their efficiency. Unlike [24], which is focused on evaluating specific ML models like CHIRP and Voting, this article takes a broader approach by considering various computational algorithms. While both studies aim to improve liver disease diagnosis, the first study specifically focuses on comparing different ML models' performances, whereas the second study delves into a wider range of computational methods and efficiency enhancement techniques for liver disease identification.

Additionally, in [24] authors evaluate the model's performance using Mean Absolute Error (MAE), Relative Absolute Error (RAE), and Accuracy metrics compared to existing models. The results indicate that CHIRP outperforms RF and MLP in terms of MAE and RAE, with an Accuracy of 71.30% which suggests the effectiveness of CHIRP for liver disease detection compared to other methods.

Recent deep learning architectures such as Transformers (Vaswani et al., 2017) have demonstrated significant potential in healthcare applications, particularly for sequential EHR data modeling, clinical text processing, and predictive diagnostics. Models like BERT and its medical variants (e.g., BioBERT, ClinicalBERT) have improved performance on clinical NLP tasks such as named entity recognition and de-identification [26]. Ethical issues remain a major concern in deploying ML in healthcare. Rajkomar et al. (2018) highlight critical challenges, including bias in training data, lack of model transparency, and disparities in care. Fairness, explainability, and data privacy should be central to ML model design, especially in high-stakes environments [27].

To conclude, this thorough literature review reveals the supremacy of machine learning in liver disease prediction. The detailed studies have demonstrated in Table 1 where the accuracies range from 83.70% to 99%. It demonstrates how effective and adaptive these techniques are in identifying liver diseases with minimal miss rate. The different algorithms used by different authors in this literature have been illustrated in Table 2.

In summary, this research shows the extent to we have come using machine learning for predicting liver disease, which can inform towards better care and outcomes through health system improvement. This represents a significant step forward in medical diagnostics, with better patient care as well as outcomes through the application of machine learning technologies. This marks a turning point in improved prognosis and treatment methods for liver diseases, making machine learning a critical tool towards enhancing diagnostic efficiency and healthcare provision. Table 1 illustrates the comparison accuracy in different literature.

Table. 1: Performance Comparison of the Recent Studies on the ILPD Dataset:

References	Title of the Work	Authors	Description	Techniques/ Algorithms used	Accuracy
[10]	Prediction of Liver Diseases Based on Machine Learning Technique for Big Data	Engy A. El-Shafeiy, Ali I. El-Desouky, and Sally M. Elghamrawy	ML models are applied to a dataset of 7,000 patients with 23 attributes to predict liver diseases, and performance is evaluated with accuracy, sensitivity, and specificity.	Naive Bayes, SVM, and Boosted C5.0	97.20%
[5]	Intelligent Identification of Liver Diseases Based on Incremental Hidden Layer Neurons ANN Model	Panduranga Vital Terlapu, Ram Prasad Reddy Sadi, Ram Kishor Pondreti, Chalapathi Rao Tippana	ML models analyze data using a Decision Tree, showing superior accuracy, and propose an Artificial Neural Network for enhanced detection.	Logistic Regression, SVM, Naïve Bayes, KNN, and Decision Tree	97.12%
[1]	Prognosis of Liver Disease: Using Machine Learning Algorithms	Vyshali J Gogi and Dr. Vijayalakshmi MN	Data Mining and Linear Discriminant algorithms diagnose liver diseases accurately, thereby enhancing medical therapy.	Data mining and Linear Discriminant Algorithm.	95.8%
[22]	Performance Analysis of Liver Disease Prediction using Machine Learning Learning Algorithms	M. Banu Priya, P. Laura Juliet, P.R. Tamils Levi	Using data mining methods, this thesis achieves a 95.04% accuracy in predicting liver disease in Indian patients.	Random forest, support vector machines	95.04%
[7]	Prediction of fatty liver disease using machine learning algorithms, Computer Methods, and Programs in Biomedicine	Chieh-Chen Wu, Wen-Chun Yeh, Wen-Ding Hsu, Md. M. Islam, Phung Anh (Alex)	Random Forest classifies fatty liver disease with an accuracy rate of 87.48%, aiding early detection of liver cirrhosis.	Random Forest, Naive Bayes, and Logistic Regression	87.48%
[18]	Liver Disease prediction using Machine Learning classification.	Sadhasivam, J., J, S., RM, G., & Chellapan, N.	Identifies patients at risk of advanced liver diseases through ML, focusing on broader clinical implications.	Support Vector Machine (SVM).	87.09%
[4]	Liver Disease Prediction and Classification using Machine Learning Techniques	Tokala, S., Hajarathaiah, Gunda, S. Nalluri, L. Nagamanohar, P. Anamalamudi, S.Enduri	Discusses factors escalating liver diseases, emphasizing the need for reliable detection methods	Logistic Regression, SVM, KNN, Random Forest	87%
[13]	A comparative analysis of machine learning algorithms to predict liver disease.	Ghosh, M., Raihan, M. M. S., Raihan, M., Akter, L., Bairagi, A. K., Alshamrani, S. S., & Masud, M	Evaluates ML algorithms for chronic liver disease prediction, emphasizing usefulness in accurate decision-making.	LR, random forest, XGBoost, support vector machine (SVM), KNN, and decision tree	83.70%

Table. 2: Different Algorithms Used in the Corresponding Papers

ML algorithm/ method	Article count	References
SVM	16	[5],[6],[8],[9],[1],[3],[4],[10],[11],[12],[13],[14],[16],[18],[21],[23]
DEEP LEARNING	2	[6],[19]
LDA	1	[1]
LR	9	[5],[6],[7],[11],[1],[2],[3],[4],[13]
DT	9	[6],[5],[9],[1],[2],[3],[11],[13],[16]
RF	11	[5],[7],[9],[2],[3],[4],[11],[12],[13],[14],[21]
KNN	5	[5],[13],[2],[3],[4]
GRADIENT BOOSTING	1	[2]
NB	6	[5],[7],[8],[3],[10],[12]
XG BOOST	3	[13],[15],[2]
SMOTE	2	[14],[23]
NEURAL NETWORK	3	[19],[21],[23]
FEATURE SELECTION	2	[14],[5]
DATA MINING	4	[6],[8],[16],[18]

We have not only summarized prior studies but also critically assessed their limitations, particularly regarding:

- Overfitting risks in models reporting very high accuracies (e.g., the 99.9% accuracy in [5]), which may stem from insufficient validation or small, homogenous datasets.
- Dataset biases, such as skewed class distributions or region-specific data (e.g., the ILPD dataset's demographic focus), can limit generalizability.
- Lack of model interpretability, where many studies focus solely on performance metrics without considering the practical applicability in clinical settings.

By highlighting these limitations, we aim to contextualize our contributions more clearly—particularly our efforts to handle class imbalance with SMOTE, apply ensemble learning for robustness, and emphasize interpretability and future incorporation of explainable AI tools.

3. General overview of the dataset and ML techniques used

3.1. Data sources and description

Building on prior ML applications, this study leverages the ILPD dataset to enhance prediction accuracy.
Having reviewed prior studies and identified key gaps in existing approaches, the next step involves an in-depth description of the dataset utilized in this study. This dataset is critical as it provides the necessary foundation for addressing the research questions and testing the hypotheses discussed in the preceding section. In this research, we collected the Indian liver patient dataset from the UCI Machine Learning Repository through Kaggle, i.e., the original dataset is from the northeast of Andhra Pradesh, India[26]. This data consists of 583 liver patients a where 75.64% are male patients and 24.36% are female patients. It consists of 10 parameters as shown in the Table.3, but we have chosen 8 parameters for our further analysis, with one parameter as our target attribute.

Table. 3: Different Parameters of Our Dataset

I.	Age: Age of the Patient
II.	Gender: Gender of the Patients
III.	TB: Total Bilirubin
IV.	DB: Direct Bilirubin
V.	Alkphos: Alkaline Phosphatase
VI.	Sgpt: Alanine Aminotransferase
VII.	Sgot: Aspartate Aminotransferase
VIII.	TP: Total Proteins
IX.	ALB: Albumin
X.	AG Ratio: Albumin and Globulin Ratio

3.2. Dataset description

We cleaned the dataset and now it has the records of 416 liver patients and 167 non-liver patients as illustrated in Fig.3.1. The "Class" column is a class label used to divide the patients into groups i.e., liver patients (having liver disease) and non-liver patients (no liver disease) out of which 441 are male patients and 142 female patients. This is shown in the gender distribution plot of Fig.3. b.
The dataset contains the columns, namely: ALB (Albumin), TB (Total Bilirubin), DB (Direct Bilirubin), TP (Total Proteins), Alkphos (Alkaline Phosphotase), Sgpt (Alanine Aminotransferase), Sgot (Aspartate Aminotransferase), A/G Ratio (Albumin and Globulin Ratio), Age, Gender, and Class.

3.3. Data preprocessing

Using mean imputation for your liver patient dataset is a practical first step—it's simple, preserves dataset size, and maintains column means—but it has important limitations, especially in clinical settings. We standardized gender labels by converting 'male' to 'M' and 'female' to 'F'. In addition to this, numerical features like ALB (Albumin), TB (Total Bilirubin), DB (Direct Bilirubin), TP (Total Proteins),

Alkphos (Alkaline Phosphatase), Sgpt (Alamine Aminotransferase), Sgot (Aspartate Aminotransferase), A/G Ratio (Albumin and Globulin Ratio) have been normalized to one scale to enable processing by machine learning algorithms. By doing this, we have been able to mitigate biases brought about by different scales of measurements, which has improved the performance of our model. Besides, age has been transformed to ensure consistency, especially for entries where age is greater than 89. Thus, any input that showed an age greater than eighty-nine has been uniformly represented as ninety. Thereafter, we refined and structured the dataset using preprocessing techniques like cleaning and encoding, leading to a strong foundation for the development of an accurate and reliable Liver Disease Prediction Model.

Fig. 3: A) Count Plot of Class.

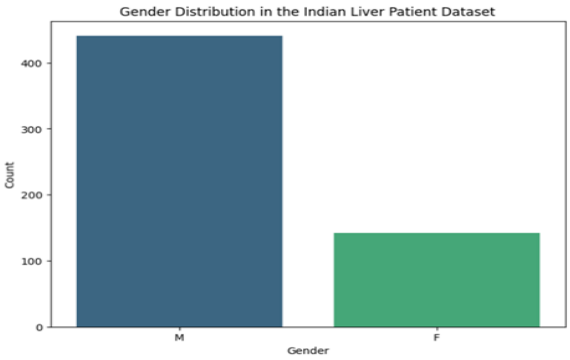

Fig. 3: B) Gender Distribution.

4. Proposed system/model

Unlike many prior studies that relied on a single classifier or simple bagging/boosting approaches, our hybrid model combines three complementary algorithms—Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF)—using a soft voting ensemble. This integration leverages their respective strengths: LR's interpretability, SVM's performance on high-dimensional data, and RF's robustness to overfitting.

4.1. Methodologies used

4.1.1. Feature selection

Feature selection refers to the process of selecting the important features from original attributes to reduce the feature space optimally based on certain criteria. This is a very important and vital step, which is shown in Fig. 4. In text classification, there may be certain words that aren't used often. For example, there is only a single document where the word "arrange" is used, and it is positive. Do we have to keep this word as a feature or not? For this, there is no simple yes or no answer since with just a single training example, we can't figure out if it goes in the positive class or if it is noise. As an alternative, we may remove it completely.

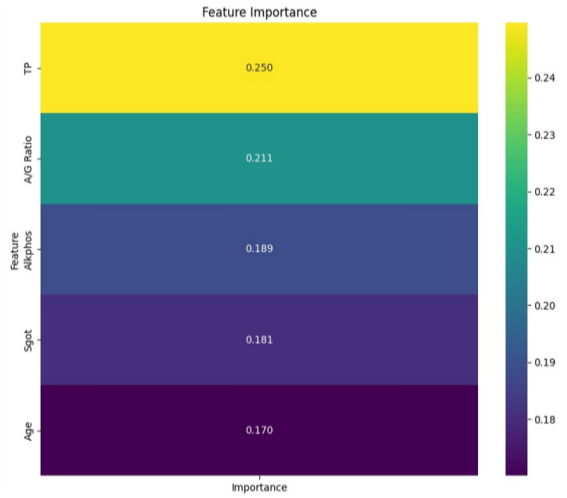

Fig. 4: Feature Selection Heatmap.

4.1.2. SMOTE

The Synthetic Minority Over-Sampling Technique (SMOTE) is applied to balance the dataset. This method increases the representation of the minority class, helping improve classifier performance on imbalanced data. SMOTE is used to treat class imbalance. There could be very few instances of some classes in a dataset (like liver disease), which creates an imbalance. Such an imbalance would result in incorrect models. This imbalance is depicted in Fig.5a. Therefore, SMOTE creates new synthetic data targeted especially for the lesser category to adjust this situation within our sets. Imagine a scenario where we possess multiple fabricated patients suffering from liver disorders for training the machine. It concentrates on generating new instances in the feature space with reference to interpolation between positive examples standing close to one another. This is illustrated in Fig.5 b, where the 'Class' column is taken as the target variable for determining whether the patient has a liver disease or not.

Fig. 5: A) Before Using SMOTE.

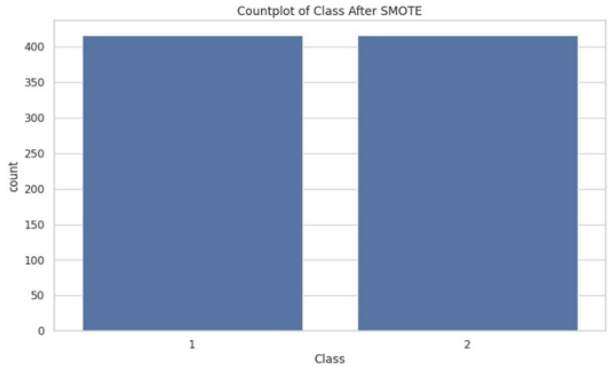

Fig. 5: B) After Using SMOTE.

It can be observed that there is a significant imbalance between the classes, with a higher number of patients having liver disease compared to those who do not. This imbalance should be taken into consideration when building and evaluating machine learning models on this dataset.

4.2. Schematic layout of the proposed model

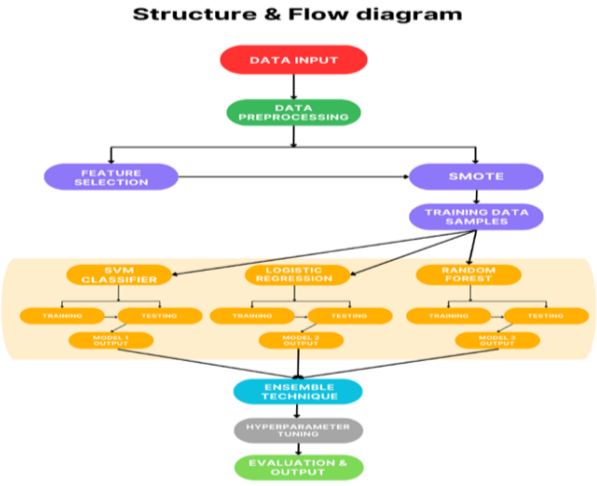

Fig. 6: Flowchart of Proposed Technique.

A detailed explanation of each phase and the techniques is shown in Fig.6.

The process of liver disease detection begins with data input, where raw data related to various attributes of patients, such as age, gender, and blood sample results, has been collected. This initial stage has set the foundation for further analysis by providing the necessary medical information.

Next, in the data preprocessing phase, the raw data has been cleaned and prepared for analysis. This has involved handling missing values, standardizing data, and encoding categorical variables to ensure compatibility with machine learning algorithms. Techniques like

normalization and encoding have been employed to make the data usable. Feature selection follows, where significant features contributing to the prediction of liver disease have been identified. This step has reduced the dimensionality of the data, enhancing model performance by focusing on the most relevant variables.

To address the class imbalance present within the dataset, we have used the concept of the Synthetic Minority Over-Sampling Technique (SMOTE). SMOTE has generated synthetic examples of the minority class, ensuring that machine learning models do not exhibit bias towards the majority class, thus improving the model's performance. The preprocessed and balanced dataset has then been split into training and test sets, typically in an 80-20 ratio, during the training data samples phase. This split has been done so that it is ensured the model is trained on a considerable chunk of the data, while a portion is left for the purpose of testing.

In the machine learning models stage, three different algorithms, i.e., SVM, RF, and LR, are employed for training and testing of the model. Each algorithm has brought its unique strengths: SVM has found the best hyperplane for classification, LR has modeled binary dependent variables using a logistic function, and RF has constructed multiple decision trees to enhance prediction stability. The outputs from these individual models have then been combined using an ensemble technique. The ensemble approach has improved overall prediction accuracy by harnessing the strengths of each ML model. Evaluation of the combined model has been performed using metrics such as accuracy, precision, recall, and F1-score. This comprehensive assessment has helped determine the model's efficiency and reliability for predicting liver disease.

Finally, the output has been generated based on the evaluation results, predicting whether liver disease is present in patients. Our model's output has been interpreted and presented in a manner that indicates whether a patient is likely to have liver disease.

This structured approach has ensured robust liver disease detection by leveraging multiple machine learning algorithms and improving prediction accuracy through ensemble techniques. By systematically collecting, preprocessing, and analyzing data, the method has provided a reliable means of identifying liver disease, aiding in early diagnosis and better clinical decision-making.

4.3. Proposed algorithm(s)

In this work, we are using three classification algorithms: SVM, LR, and RF. SVM is employed due to its effectiveness in handling high-dimensional and non-linear data scenarios. LR is particularly effective for binary classifications and linearly separable data. RF, being an ensemble method, builds decisions based on multiple DTs on various subsets of data, then integrates their outcomes for enhanced accuracy and to prevent overfitting. Here is a further description of each algorithm, which we have inculcated in this work

4.3.1. Logistic regression

Logistic Regression is an algorithm that is used mainly for binary classification. It is a statistical method that is used to predict the probability of a dependent variable based on independent variables. It consists of a logistic function that generates output in the range 0 and 1, which thereby makes it useful for predicting binary outcomes like Yes/No or True/False. It uses a Logistic function to estimate the relation between the factors that can impact the outcome.

The Logistic Regression equation is given as, $p = (f(x) = \frac{1}{1 + e^{-x}})$

Logistic Regression's Fundamental Benefits:

- Easy machine learning techniques:
- Applicable for linearly separable data sets
- Valuable inferences

4.3.2. Random forest (RF)

An ensemble of random forests is an ML technique used for classification and regression that works by the construction of an enormous number of decision trees during the training time and results in the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The importance of different features using a random forest is shown in Fig.7. The propensity of decision trees to overfit their training set makes them inappropriate for a random forest. For example, among the forest's trees, there may be some that consider only one variable while others consider many variables at a time, as well as their interactions with each other. To obtain more efficient and accurate predictions, additional randomness is introduced into the random forest.

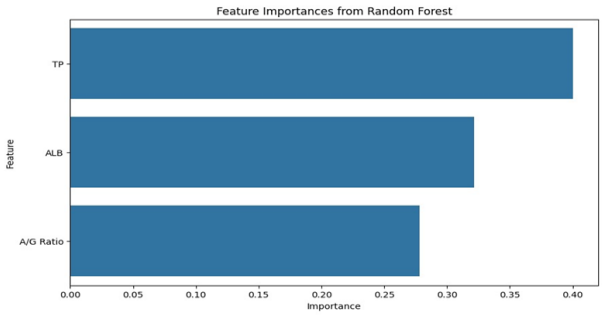

Fig. 7: Feature Importance of Using Random Forest.

4.3.3. Support vector machine (SVM)

Support Vector Machine (SVM) is a supervised ML algorithm that can be used for classification as well as regression. In SVM, the goal is to create a decision boundary, also known as a Hyperplane, that can be used to maximize the boundary between different classes. It can be used especially for separating high-dimensional data and is even useful in scenarios where the data is not linearly separable.

5. Experimentation and model evaluation

5.1. Depiction of results

This section throws light on the findings of the experimental study, which has been discussed in detail. The libraries used include pandas, which have been employed for data manipulation and preprocessing tasks such as reading the dataset and handling missing values. Scikit-learn has been utilized for preprocessing, building machine learning models, hyperparameter tuning, and evaluation. Imbalanced-learn has applied SMOTE for handling class imbalance in the training data, while NumPy has been used for numerical operations and array manipulation. The Voting Classifier has combined the predictions of the SVM, Logistic Regression, and Random Forest classifiers using soft voting for the final ensemble model. Grid Search Cross Validation has performed hyperparameter tuning for each classifier to find the best model parameters, and SMOTE has resampled the training data for balancing the class distribution. The visualization of the correlation heatmap is shown in Fig. 8. The Correlation heatmap shows pairwise relationships among features. Strong positive or negative correlations highlight potential multicollinearity or redundant features, informing feature selection decisions.

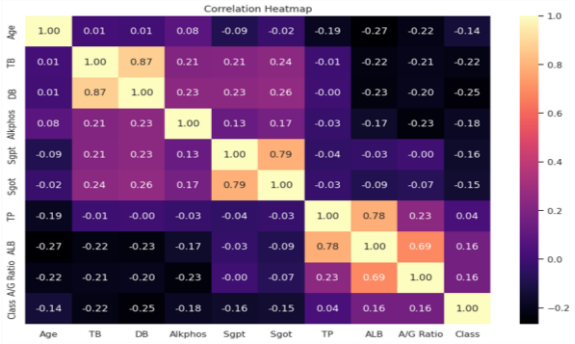

Fig. 8: Correlation Heatmap of All the Features.

5.2. Validation/system performance evaluation

In the context of evaluating classification algorithms, particularly for medical diagnoses such as liver disease prediction, it is crucial to understand and utilize various evaluation metrics. The metrics discussed here are derived from the confusion matrix, which categorizes predictions into True positives (TP), False positives (FP), True negatives (TN) as well and False negatives (FN). Recall refers to the measurement of the ability of the classifier to correctly identify patients with liver disease. High recall is crucial in medical diagnostics to ensure that as many positive cases as possible are identified, thus minimizing the number of false negatives (missed diagnoses). The F1 score is the harmonic mean of precision and recall. The F1 score provides a single metric that balances both precision and recall, making it particularly useful when the classes are imbalanced. A high value of F1 score means that the classifier has both high precision as well as high recall, making it a reliable measure of the model's overall performance. The various performance metrics of different algorithms are recorded in Table 4, and in Table 5, 10-fold cross-validation of the hybrid model is shown. Performance metrics (accuracy, precision, recall, F1-score) for each model were evaluated. These results compare classification effectiveness and highlight the best-performing algorithm. We compared the accuracy, precision, recall, and F1 score for each model and mentioned them in tables.

Table 4: Comparison of Various Performance Metrics on Various Machine Learning Models

Model	Precision	Recall	F1 score	Accuracy
Random Forest	0.77	0.82	0.79	0.6842 ~ 68%
SVM	0.72	0.94	0.82	0.6930 ~ 69%
Logistic Regression (LR)	0.74	0.96	0.84	0.7193 ~ 72%

Table 5: 10-Fold Cross-Validation of Hybrid Model

N (N = 10) Fold Cross Validation									
1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	6-Fold	7-Fold	8-Fold	9-Fold	10-Fold
0.9787	0.9893	1	0.9677	0.9784	1	0.9892	0.9784	1	1

Table 6: Table Illustrating Model Accuracy

Model	Accuracy
Ensemble model	92.30%
Mean cross-validation	98.82%
Final model	98.8%

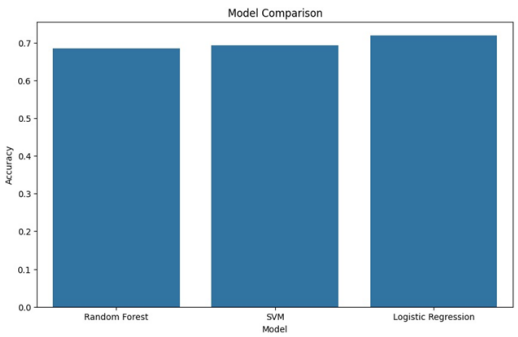

Fig. 9: Model Comparison.

5.3. Discussions and analysis

Various classifiers, in addition to an ensemble voting classifier have been used to implement the following model. The dataset, "indian_liver_patient.csv," has gone through several preprocessing steps such as removing the 'Gender' column, mean normalization of missing values, and converting the target variable 'ALB' (continuous) into discrete classes using binning and encoding. Numerical features were processed differently from categorical features. While imputation was used for numerical ones, one-hot encoding was employed for the categorical ones.

Afterward, the data was divided into training and testing sets, where SMOTE balanced class imbalance on the training set. Grid Search Cross Validation to identify the best hyperparameters helped fine-tune three classifiers i.e., Support Vector Machine (SVM), Logistic Regression (LR), and Random Forests (RF). Each of them is passed on to a Voting Classifier, which combines their predictions using the soft voting method. 10-fold cross-validation and accuracy on the test set were used to assess model performance, which is shown in Table 6. Fig. 9 depicts the comparison of the mentioned model's accuracy.

Among key libraries that have been used in this model include pandas for data manipulation, sci-kit-learn for preprocessing, model training, as well as evaluation, imbalanced-learn for SMOTE, and NumPy for numerical operations. The code's approach has ensured robust model training and evaluation, leveraging ensemble learning for potentially improved predictive performance.

Ethical implications: We have expanded the "Conclusion and Future Directions" in the next section to address ethical considerations such as algorithmic bias, interpretability, and data privacy. We also mention the importance of explainable AI tools (e.g., LIME, SHAP) and regulatory compliance for clinical adoption.

6. Conclusion and future directions

Many ensemble models in previous works did not explicitly manage skewed datasets. We incorporate SMOTE to synthetically balance the classes before model training, improving generalization and fairness. Classification techniques have been used in this work for efficient as well as early diagnosis of Liver disease. Algorithms like SVM, Random Forest, and LR are implemented, and their results are integrated using the Ensemble method. The output is tested based on evaluation metrics like precision, recall, accuracy, and F1-score. The accuracy of the ensemble model is found to be 92.30% and after applying 10-Fold Cross Validation, the mean cross-validation accuracy is found to be 98.82%. The hybrid architecture improves prediction accuracy for liver disease. While some studies report high accuracy without cross-validation (which may mask overfitting), our model is evaluated using 10-fold cross-validation, ensuring robustness and minimizing performance inflation. We emphasize:

- Class imbalance: Unlike several prior studies that overlooked this issue, our model applies SMOTE to ensure balanced representation and improved predictive fairness.
- Model interpretability: While many high-performing models (e.g., deep neural networks) sacrifice explainability, we selected interpretable algorithms (e.g., LR, RF) and propose future integration of LIME/SHAP to enhance clinical trust and usability.
- Realistic performance expectations: We contextualize our results by acknowledging overfitting risks in prior studies with unusually high reported accuracies, presenting our ensemble model's robust validation (e.g., 10-fold cross-validation) as a more grounded approach.

The relevance of deep learning models such as CNNs and Transformer-based architectures, as well as the application of generative AI (e.g., GANs) for augmenting clinical datasets. These additions point toward promising directions for improving both the accuracy and robustness of predictive models.

To enhance the robustness and applicability of the proposed model, future research could focus on validating its performance across diverse and real-world clinical datasets, including those from different demographics, geographic regions, and healthcare systems. This would help assess generalizability and uncover potential biases. Additionally, integrating more advanced deep learning architectures—such as convolutional or transformer-based models—may further improve predictive accuracy, particularly for unstructured data like medical images or clinical notes. Finally, the incorporation of interpretability tools such as LIME (Local Interpretable Model-agnostic Explanations) or SHAP (Shapley Additive exPlanations) could help provide transparent insights into model decision-making, increasing clinician trust and facilitating regulatory compliance in healthcare AI deployments.

7. Acknowledgement

The authors express sincere thanks to all reviewers whose valuable feedback and suggestions have significantly improved the quality of this work. Appreciation is also extended to SOA University for the resources that provided the necessary support during the preparation of this article.

References

[1] Gogi, V. J., & M. N., V. (2018). Prognosis of liver disease: Using machine learning algorithms. 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 875–879. https://doi.org/10.1109/ICRIEECE44171.2018.9008482.

[2] Gupta, K., Jiwani, N., Afreen, N., & D, D. (2022). Liver disease prediction using ML classification techniques. 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 221–226. https://doi.org/10.1109/CSNT54456.2022.9787574.

[3] Rahman, A. K. M. S., Shamrat, F. M. J. M., Tasnim, Z., Roy, J., & Hossain, S. A. (2019). A comparative study on liver disease prediction using supervised machine learning algorithms. International Journal of Scientific & Technology Research, 8(11), 419–422.

[4] Terlapu, P., Sadi, R., Pondreti, R., & Tippana, C. (2022). Intelligent identification of liver diseases based on an incremental hidden layer neurons ANN model. International Journal of Computing and Digital Systems, 11(1), 1027–1050. https://doi.org/10.12785/ijcds/110183.

[5] Yadav, E. S., Bachepalli, S., & Prasanna, K. (2023). Liver disease prediction using machine learning. International Journal of Advance Research and Innovative Ideas in Education, 9(2), 65–69.

[6] Wu, C.-C., Yeh, W.-C., Hsu, W.-D., Islam, M. M., Nguyen, P. A., Poly, T. N., Wang, Y.-C., Yang, H.-C., & Li, Y.-C. (2019). Prediction of fatty liver disease using machine learning algorithms. Computer Methods and Programs in Biomedicine, 170, 23–29. https://doi.org/10.1016/j.cmpb.2018.12.032.

[7] Amin, R., Yasmin, R., Ruhi, S., Rahman, M. H., & Reza, M. S. (2023). Prediction of chronic liver disease patients using integrated projection-based statistical feature extraction with machine learning algorithms. Informatics in Medicine Unlocked, 36, 101155. https://doi.org/10.1016/j.imu.2022.101155.
[8] Spann, A., Yasodhara, A., Kang, J., Watt, K., Wang, B., Goldenberg, A., & Bhat, M. (2020). Applying machine learning in liver disease and transplantation: A comprehensive review. Hepatology, 71(3), 1093–1105. https://doi.org/10.1002/hep.31103.
[9] El-Shafeiy, E., El-Desouky, A., & Elghamrawy, S. (2018). Prediction of liver diseases based on machine learning technique for big data. In Advanced Machine Learning Technologies and Applications (AMLTA2018) (Vol. 723). Springer. https://doi.org/10.1007/978-3-319-74690-6_36.
[10] Razmpour, F., Daryabeygi-Khotbehsara, R., Soleimani, D., & others. (2023). Application of machine learning in predicting non-alcoholic fatty liver disease using anthropometric and body composition indices. Scientific Reports, 13, 4942. https://doi.org/10.1038/s41598-023-32129-y.
[11] Ahn, J. C., Connell, A., Simonetto, D. A., Hughes, C., & Shah, V. H. (2021). Application of artificial intelligence for the diagnosis and treatment of liver diseases. Hepatology, 73(6), 2546–2563. https://doi.org/10.1002/hep.31603.
[12] Ghosh, M., Raihan, M. M. S., Raihan, M., Akter, L., Bairagi, A. K., Alshamrani, S. S., & Masud, M. (2021). A comparative analysis of machine learning algorithms to predict liver disease. Intelligent Automation & Soft Computing, 30(3), 917–928. https://doi.org/10.32604/iasc.2021.017989.
[13] Aslam, M. H., Hussain, S. F., & Ali, R. H. (2022). Predictive analysis on severity of non-alcoholic fatty liver disease (NAFLD) using machine learning algorithms. 17th International Conference on Emerging Technologies (ICET), 95–100. https://doi.org/10.1109/ICET56601.2022.10004660.
[14] Liu, Y., Liu, X., Cen, C., Li, X., Liu, J., Ming, Z., Yu, S., Tang, X., Zhou, L., Yu, J., Huang, K., & Zheng, S. (2021). Comparison and development of advanced machine learning tools to predict non-alcoholic fatty liver disease: An extended study. Hepatobiliary & Pancreatic Diseases International, 20(5), 409–415. https://doi.org/10.1016/j.hbpd.2021.08.004.
[15] Perveen, S., Shahbaz, M., Keshavjee, K., & others. (2018). A systematic machine learning-based approach for the diagnosis of non-alcoholic fatty liver disease risk and progression. Scientific Reports, 8, 2112. https://doi.org/10.1038/s41598-018-20166-x.
[16] Ghazal, T. M., Rehman, A. U., Saleem, M., Ahmad, M., Ahmad, S., & Mehmood, F. (2022). Intelligent model to predict early liver disease using machine learning technique. International Conference on Business Analytics for Technology and Security (ICBATS), 1–5. https://doi.org/10.1109/ICBATS54253.2022.9758929.
[17] Sadhasivam, J., J, S., RM, G., & Chellapan, N. (2021). Liver disease prediction using machine learning classification. Webology, 18(2), 441–452. https://doi.org/10.14704/WEB/V18SI02/WEB18293.
[18] Zhou, L. Q., Wang, J. Y., Yu, S. Y., Wu, G. G., Wei, Q., Deng, Y. B., Wu, X. L., Cui, X. W., & Dietrich, C. F. (2019). Artificial intelligence in medical imaging of the liver. World Journal of Gastroenterology, 25(6), 672–682. https://doi.org/10.3748/wjg.v25.i6.672.
[19] Md, A. Q., Kulkarni, S., Joshua, C. J., Vaichole, T., Mohan, S., & Iwendi, C. (2023). Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. Biomedicines, 11, 581. https://doi.org/10.3390/biomedicines11020581.
[20] Mostafa, F., Hasan, E., Williamson, M., & Brock, M. N. (2021). Statistical ML approaches to liver disease prediction. Livers, 1(4), 294–312. https://doi.org/10.3390/livers1040023.
[21] Priya, M. B., Juliet, P. L., & Tamilselvi, P. R. (2018). Performance analysis of liver disease prediction using machine learning algorithms. International Research Journal of Engineering and Technology, 5(1), 206–211.
[22] Dristas, E., & Trigka, M. (2023). Supervised ML models for liver disease risk prediction. Computers, 12(1), 19. https://doi.org/10.3390/computers12010019.
[23] Khan, B., Naseem, R., Ali, M., Arshad, M., & Jan, N. (2019, November 23). Machine learning approaches for liver disease diagnosing. International Journal of Data Science and Analytics Applications. https://ijdsaa.com/index.php/ welcome/article/ view/71. https://doi.org/10.69511/ijdsaa.v1i1.71.
[24] Rani, S., & Dey, N. (2022). Machine learning in healthcare: Challenges and opportunities. In N. Dey, A. S. Ashour, & S. Balas (Eds.), Machine Learning in Biomedical Science (Chap. 18). Wiley. https://doi.org/10.1002/9781119785620.ch18.
[25] Ramana, B., Surendra, M., Prasad Babu, G., & Nagasuri, B. V. (2012). A critical comparative study of liver patients from USA and India: An exploratory analysis. International Journal of Computer Science, 9.
[26] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
[27] Rajkomar, A., Dean, J., & Kohane, I. (2018). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259.