RESEARCH PAPER
Contrib. Sci. & Tech Eng, 2024, 1(1)
Homepage: https://cste.journals.umz.ac.ir/
DOI: 10.22080/cste.2024.5012

Ensemble-Based Detection and Classification of Liver Diseases Caused by Hepatitis C

Hannah Yousefpour [1], Jamal Ghasemi [1]*

[1] *Faculty of Engineering & Technology, University of Mazandaran, Babolsar, Iran.*

Article Info
Received 01 January 2024
Accepted 15 February 2024
Available online 15 March 2024

Keywords:
Liver Diseases;
Machine Learning Algorithms;
LightGBM;
AdaBoost;
Random Forest;
XGBoost.

Abstract:
The liver, as the largest internal organ in the human body, plays a pivotal role in numerous physiological processes, orchestrating over 500 metabolic activities crucial for maintaining bodily functions. However, the Hepatitis C Virus (HCV) poses a grave threat to liver health, necessitating early identification of liver diseases to halt the progression to carcinoma and potentially save lives. This research aims to train ensemble-based algorithms for classifying and detecting Hepatitis, Fibrosis, and Cirrhosis. Employing rigorous preprocessing techniques, 80% of the dataset was allocated to train five ensemble-based algorithms: AdaBoost, Random Forest, Rotation Forest, XGBoost, and LightGBM. These algorithms were evaluated across four performance metrics—accuracy, precision, recall, and F1-score. Remarkably, LightGBM emerged as the frontrunner, boasting an exceptional accuracy rate of 98.37%. Rotation Forest followed closely with an accuracy of 96.74%, while XGBoost attained an accuracy of 95.12%. Random Forest and AdaBoost secured 94.19% and 93.30% accuracy, respectively. These findings underscore LightGBM's prowess as a promising algorithm for detecting and classifying liver diseases. By leveraging advanced machine learning techniques, particularly ensemble-based algorithms, this research contributes to the ongoing efforts to enhance early detection, improve patient outcomes, and foster more effective management strategies for liver-related ailments in clinical settings.

***Corresponding Author:** j.ghasemi@umz.ac.ir

Supplementary information: Supplementary information for this article is available at https://cste.journals.umz.ac.ir/

Please cite this paper as: Yousefpour, H., & Ghasemi, J. (2024). Ensemble-Based Detection and Classification of Liver Diseases Caused by Hepatitis C. Contributions of Science and Technology for Engineering, 1(1), 32-42. doi:10.22080/cste.2024.5012.

1. Introduction

The liver is the largest organ inside the body, accounting for approximately 2% to 3% of average body weight. It is located in the upper right portion of the abdomen. A liver cell plays a crucial role in various physiological processes by performing over 500 metabolic activities, which makes it one of the most complex organs in the body [1]. Additionally, the liver is responsible for storing glucose in the form of glycogen and releasing it into the bloodstream as needed to maintain stable blood sugar levels, which provide energy to the body between meals or during periods of increased energy demand [2]. The liver also helps to maintain homeostasis by adding nutrients to the blood.

The liver is involved in the synthesis of various proteins essential for blood clotting, immune function, and maintaining fluid balance within the body. It also plays a crucial role in detoxifying the blood by removing harmful substances such as toxins, drugs, and alcohol and metabolizing these substances into less harmful byproducts that can be excreted through urine or bile [3]. The liver also produces bile, which is important in the mechanical digestion of fats. One of the most important functions of the liver is the production of bile, a digestive fluid essential for the breakdown and absorption of fats in the small intestine. Bile is stored in the gallbladder and released into the small intestine to aid digestion and absorption of dietary fats [4].

Upon entry into the body, HCV primarily infects hepatocytes, the main functional cells of the liver. The virus hijacks the host cell's machinery to replicate itself, producing numerous viral particles. This replication process often triggers an immune response, resulting in inflammation within the liver [4]. As the infection progresses, the continuous inflammation and damage to liver cells can lead to the development of fibrosis, a process in which scar tissue replaces healthy liver tissue. Persistent fibrosis can progress to cirrhosis, a condition characterized by extensive scarring and disruption of liver function. Cirrhosis significantly impairs the liver's ability to perform essential tasks, such as detoxification, metabolism, and production of vital proteins [2].

HCV has the ability to evade the immune system and establish chronic infection in the liver. The virus can evade detection by the immune system through various mechanisms, including mutations in its genetic material and interference with immune cell function. Chronic HCV

*ISSN **3060-6578***

infection can persist for years or even decades, leading to ongoing liver inflammation and progressive liver damage [5]. Furthermore, HCV can directly damage liver cells by inducing oxidative stress and promoting fibrosis, the formation of scar tissue in the liver. Oxidative stress occurs when there is an imbalance between the production of reactive oxygen species (ROS) and the body's ability to neutralize them. ROS can damage cellular components, including proteins, lipids, and DNA, leading to cell injury and death [6]. Moreover, chronic HCV infection is a major risk factor for the development of hepatocellular carcinoma (HCC), the most common type of liver cancer. The mechanisms underlying HCV-associated liver cancer are complex and multifactorial, involving both direct viral effects and indirect pathways related to chronic inflammation, fibrosis, and impaired liver regeneration [7]. Figure 1 represents the process of a healthy liver getting infected with HCV to the cancer stage.

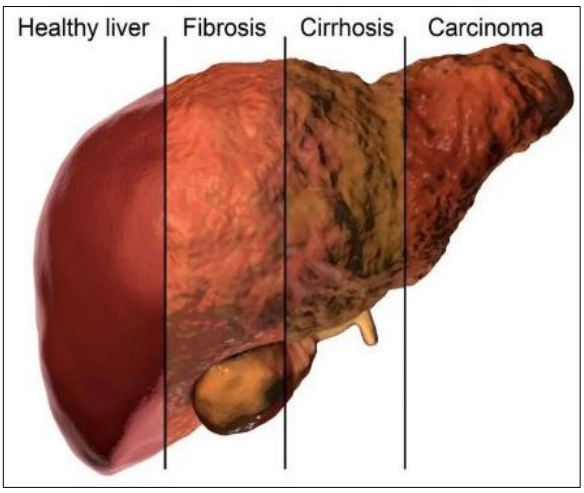


Figure 1. Illustration of the progression of hepatitis C infection in the liver [38]

Early identification of liver disease can potentially save lives and prevent the progression of hepatitis C to carcinoma. Detecting liver ailments at an early stage can significantly improve patients' longevity [2]. Numerous studies have been conducted to diagnose liver abnormalities and classify liver conditions. These studies have utilized individual algorithms as well as ensemble algorithms, with some focusing on integrating such algorithms. This study aims to train ensemble-based algorithms to classify and detect Hepatitis, Fibrosis, and Cirrhosis. Ensemble-based algorithms are known for their efficiency compared to other machine learning algorithms. The present research is structured into five sections: the first section introduces the importance of the liver in the body and emphasizes the significance of early diagnosis; the second section comprehensively reviews existing literature on liver diagnosis research; the third section discusses the dataset, analyses data, and explores ensemble-based algorithms; the fourth section presents the results obtained from the application of ensemble learners; and finally, the last section provides a conclusion summarizing the study's findings.

2. Literature Review

Harabor et al. [8] investigated the use of machine learning algorithms, including support vector machines (SVM) and random forests, for the diagnosis of hepatitis C virus (HCV) infection. This study used a dataset comprising clinical parameters such as liver enzyme levels, viral load, and patient demographics. Results indicated that machine learning models achieved high accuracy in discriminating between HCV-infected and non-infected individuals. Zhang et al. [9] worked on developing a predictive model for assessing the risk of liver fibrosis progression in patients with chronic hepatitis C using machine learning algorithms. They used a combination of clinical, laboratory, and imaging data to train the predictive model in their study and revealed that machine learning-based risk prediction models could accurately classify patients. Elshewey et al. [10] explored a machine-learning algorithm for Optimizing HCV disease prediction in Egypt, using UCI machine-learning clinical data to train the proposed method predictive model named hyOPTGB framework. This study has been compared to other machine learning algorithms such as decision tree (DT), SVM, dummy classifier (DC), ridge classifier (RC), and bagging classifier (BC), showing great potential for prediction.

Feng et al. [11] studied the application of machine learning techniques such as deep learning and machine learning in the early detection of HCC among patients with chronic hepatitis C. The researchers developed a predictive model using a combination of clinical, imaging, and genetic data to identify individuals at high risk of developing HCC. Arjmand et al. [12] explored the application of deep learning techniques for the automated detection of HCV infection using histopathological images of liver tissue samples. The study developed a convolutional neural network (CNN) model trained on a large dataset of liver biopsy images to classify tissue samples as either HCV-infected or non-infected. Lu et al. [13] employed machine learning algorithms to predict treatment response in patients with chronic hepatitis C undergoing direct-acting antiviral therapy. Machine learning techniques such as logistic regression and decision trees were utilized to predict treatment outcomes, including sustained virological response (SVR). The findings underscored the potential of machine learning-based prediction.

Park et al. [14] investigated the use of machine learning algorithms, including Feedforward Neural Networks (FNN) and random forests, to diagnose HCV infection. This study used a dataset comprising clinical parameters such as liver enzyme levels, viral load, and patient demographics. Results indicated that machine learning models achieved high accuracy in discriminating between HCV-infected and non-infected individuals. Akella and Akella [15] worked on developing a predictive model for assessing the risk of liver fibrosis progression in patients with chronic hepatitis C using machine learning algorithms. A combination of clinical, laboratory, and imaging data was used to train the predictive model in this study, revealing that machine learning-based risk prediction models could accurately classify patients. Maiellaro et al. [16] used machine learning algorithms to predict treatment responses and outcomes in patients with hepatitis C infection. They used longitudinal

clinical data, including viral load measurements, treatment regimens, and patient demographics, to train predictive models. Neural networks were employed to forecast treatment response and identify factors influencing treatment outcomes, showing great potential for prediction.

Edeh et al. [17] studied the application of machine learning techniques, such as deep learning and ensemble learning, in the early detection of HCC among patients with chronic hepatitis C. The researchers developed a predictive model using a combination of clinical, imaging, and genetic data to identify individuals at high risk of developing HCC. Prakash et al. [18] explored the application of deep learning techniques for the automated detection of HCV infection using histopathological images of liver tissue samples. The study developed a CNN model trained on a large dataset of liver biopsy images to classify tissue samples as either HCV-infected or non-infected. Park et al. [19] employed machine learning algorithms to predict treatment response in patients with chronic hepatitis C undergoing direct-acting antiviral therapy. Machine learning techniques such as logistic regression and decision trees were utilized to predict treatment outcomes, including sustained virological response. The findings underscored the potential of machine learning-based prediction.

Wang et al. [20] used machine learning algorithms such as feature selection and classification algorithms to identify potential biomarkers associated with hepatitis C-related liver fibrosis progression. The researchers analysed multi-omics data from patients with chronic hepatitis C and varying degrees of liver fibrosis. Butt et al. [21] explored the role of six machine learning algorithms in predicting the risk of HCC development in patients with chronic hepatitis C. Machine learning algorithms, including random forests, logistic regression (LR), SVM, K-nearest neighbors (KNN), decision tree (DT), random forest (RF), and adaptive boosting (Adaboost), were employed to analyse and integrate multi-dimensional data for accurate risk prediction. The outcome underscored the potential of machine learning-based risk prediction models in identifying early intervention to prevent HCC development. Zhang et al. [22] proposed a new deep-learning model using ANN, specifically Artificial Back-Propagation Neural Network, to classify different stages of liver fibrosis in patients with chronic hepatitis C using non-invasive biomarkers. The study showed the potential of machine learning-based approaches in non-invasive fibrosis assessment, offering an alternative to liver biopsy for disease staging in hepatitis C patients.

Lilhore et al. [23] proposed a novel approach by developing a Hybrid Predictive Model (HPM) based on an improved random forest algorithm and SVM. The HPM integrates the strengths of both algorithms to enhance diagnostic accuracy and overcome the shortcomings of existing ML-based prediction models. The proposed HPM successfully improved performance in HCV diagnosis. Alizargar et al. [24] studied various machine learning techniques for predicting and early diagnosis of liver disease. Six machine learning algorithms, including Support Vector Machine, K-nearest Neighbors, SVM, Gaussian Naïve Bayes, decision tree, random forest, logistic regression, and K-nearest neighbors (KNN), Logistic Regression, decision tree, XGBoost, and ANN, were applied to two datasets. This study contributes to the advancement of predictive analytics in healthcare, offering valuable insights for early disease detection and improved patient care. Safdari et al. [25] tested six classification models. These models encompassed the support vector machine, Gaussian Naïve Bayes, decision tree, random forest, logistic regression, and KNN algorithms. The classifiers were implemented using the Python programming language. Model performance evaluation involved receiver operating characteristic curve analysis and other relevant metrics.

Overall, machine learning algorithms are promising to enhance Hepatitis C diagnosis and management. By leveraging diverse datasets and advanced computational techniques, these models can provide valuable insights into disease progression, ultimately contributing to improved clinical decision-making and patient care. In this age, machine learning algorithms offer valuable insight for healthcare professionals in their efforts to combat this infectious disease.

3. Methodology

In this section, the dataset is introduced and visualized using pair plots and correlation matrix. Pair plots provide a comprehensive overview of the relationships between variables in the dataset. After exploring the dataset, five ensemble-based algorithms, including AdaBoost, Random Forest, Rotation Forest, XGBoost, and LightGBM, were chosen for their effectiveness in handling complex datasets and their ability to generate robust predictions by aggregating multiple weak learners. Eventually, four performance metrics were introduced and used to evaluate the performance of these ensemble-based algorithms. These metrics included accuracy, precision, recall, and F1 score. The flowchart of the proposed method is represented in Figure 2.

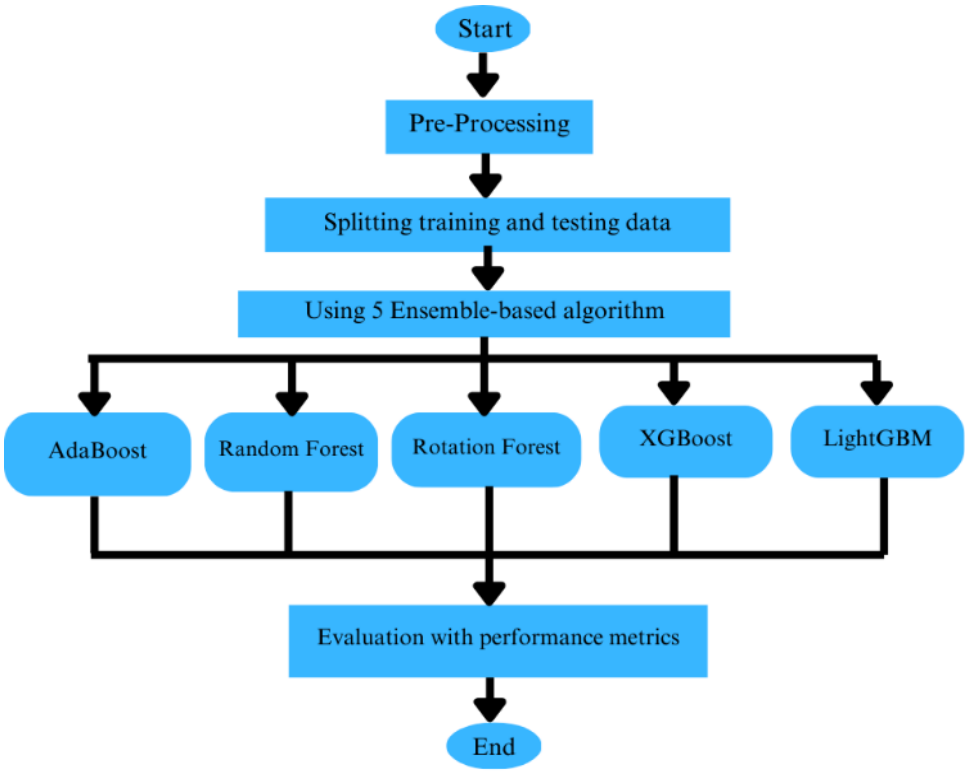


Figure 2. **Proposed method flowchart**

3.1. Dataset

HCV dataset [26] was used in this research, containing information on 615 patients and 12 features related to Hepatitis C virus (HCV) infection and eventually four statuses to classify the liver status. Table 1 represents dataset's feature description of features such as Albumin (ALB), Alkaline Phosphatase (ALP), Alanine Aminotransferase (v), Aspartate Aminotransferase (AST), Bilirubin (BIL), Serum Cholinesterase (CHE), Cholesterol (CHOL), Creatinine (CREA), Gamma-Glutamyl Transferase (GGT), Total Protein (PROT), age, and gender.

Table 1. **Dataset statistics**

Feature	Count	Mean	Std	Min	25%	50%	75%	Max
Age	615.0	47.408130	10.055105	19.00	39.000	47.00	54.000	77.00
Gender	615.0	1.386992	0.487458	1.00	1.000	1.00	2.000	2.00
ALB	615.0	41.620195	5.775920	14.90	38.800	41.90	45.200	82.20
ALP	615.0	68.283920	25.643955	11.30	52.950	66.70	79.300	416.60
ALT	615.0	28.450814	25.448940	0.90	16.400	23.00	33.050	325.30
AST	615.0	34.786341	33.090690	10.60	21.600	25.90	32.900	324.00
BIL	615.0	11.396748	19.673150	0.80	5.300	7.30	11.200	254.00
CHE	615.0	8.196634	2.205657	1.42	6.935	8.26	9.590	16.41
CHOL	615.0	5.368099	1.123466	1.43	4.620	5.31	6.055	9.67
CREA	615.0	81.287805	49.756166	8.00	67.000	77.00	88.000	1079.10
GGT	615.0	39.533171	54.661071	4.50	15.700	23.30	40.200	650.90
PROT	615.0	72.044137	5.398234	44.80	69.300	72.20	75.400	90.00

Correlation matrix provide insights into the relationships between variables, helping analysts understand how changes in one variable may affect another [27]. A correlation matrix displays the correlation coefficients between features in the dataset. Each cell in the table represents the correlation between two variables, with values ranging from -1 to 1. A correlation coefficient close to 1 indicates a strong positive relationship, while a coefficient close to -1 indicates a strong negative relationship. Figure 3 represents the correlation matrix of the used dataset in this research. Figure 4 represents a density plot of the features.

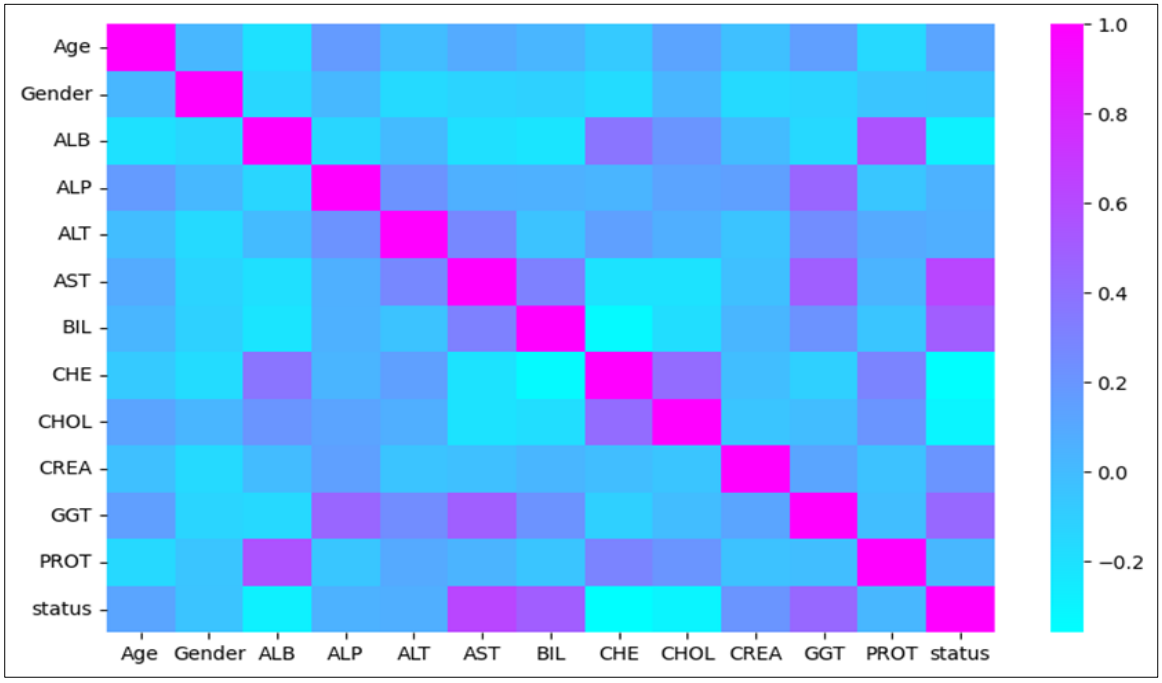
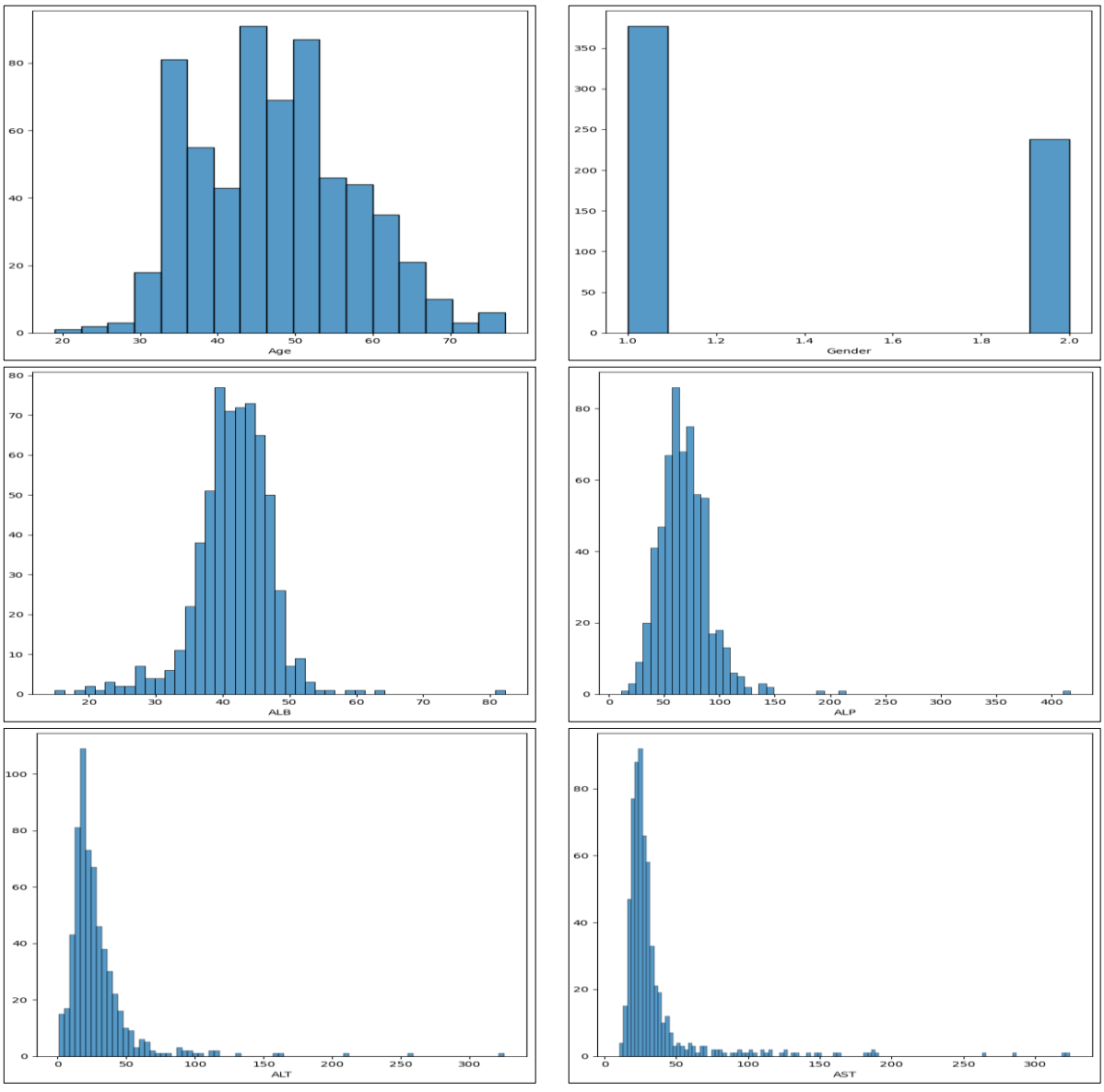


Figure 3. **Correlation matrix of the features**



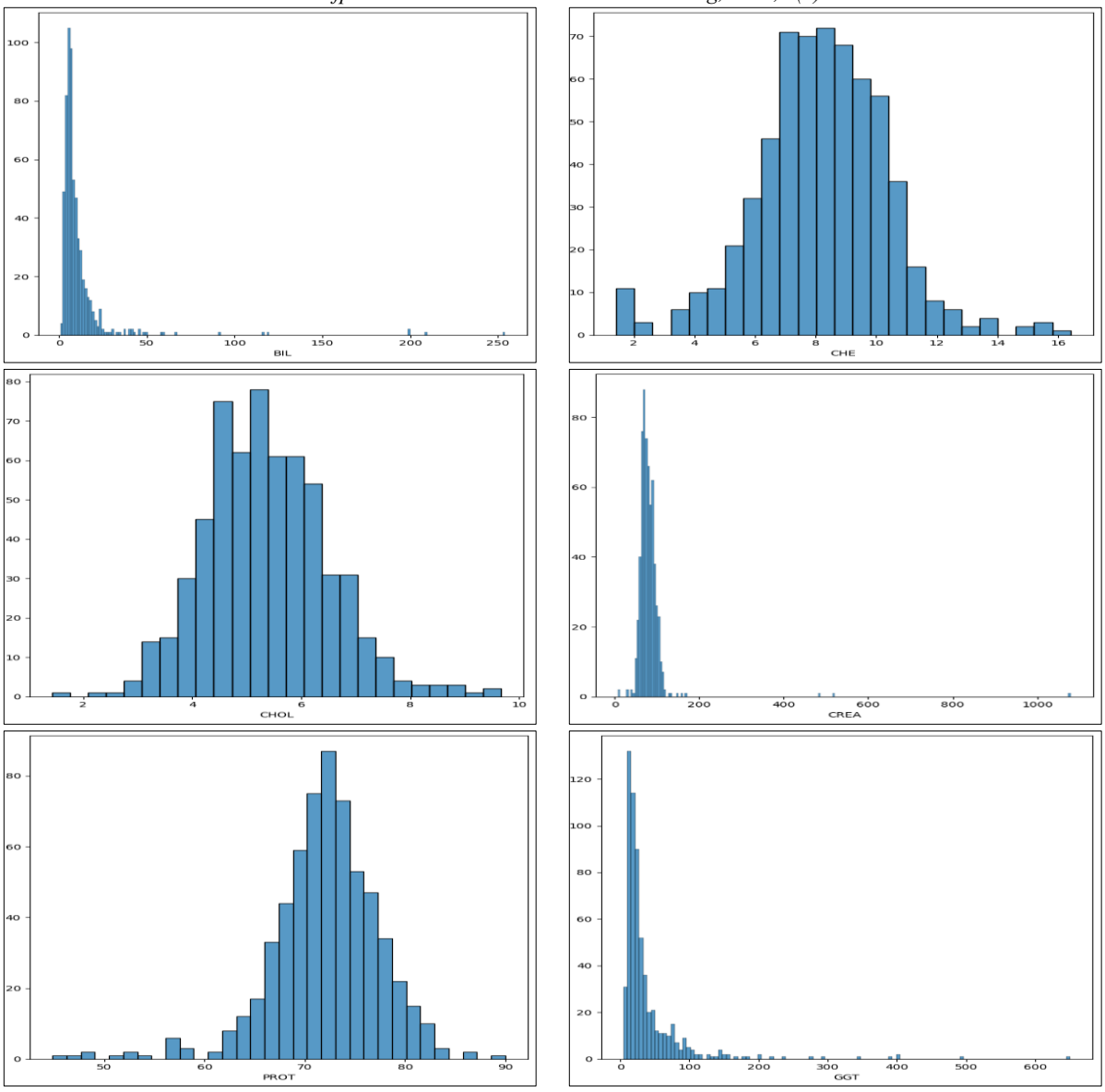


Figure 4. **Features density plot**

Pair plots, which show pairwise relationships between variables in a dataset, can also provide insightful information. In a pair plot, each variable in the dataset is plotted against every other variable, resulting in a grid of scatterplots. Along the diagonal of the grid, histograms or kernel density estimates are typically shown to visualize the distribution of each variable [28]. Figure 5 represents a pair plot of the features.

3.2. Pre-processing

In this research, pre-processing was applied before training the models. The pre-processing steps involved two main operations: replacing missing values (NaN values) with the mean of the existing values and normalizing the data. In this first step, missing values were replaced with the mean of the existing values in the same feature column. Replacement of necessary braces missing values can adversely affect the performance of machine learning models if left unaddressed. After replacing missing values, the next step in pre-processing was to normalize the data.

Normalization is a scaling technique used to standardize the range of features in the dataset. By normalizing the data, values of each feature are set to a similar range, typically between -1 and 1. This ensures that all features contribute equally to the model's training process and prevents features with larger scales from dominating the learning process.

3.3. Ensemble Machine Learning Algorithms

This section delves into the detailed explanation of the five ensemble-based machine learning algorithms employed in this research. Ensemble learning techniques are powerful methodologies that use multiple models to improve predictive performance and robustness. The ensemble methods utilized in this study have been chosen for their versatility, effectiveness, and widespread application in various domains, as ensemble-based algorithms have performed better than individual learners. Table 2 represents the algorithm setting of the algorithms' parameters.

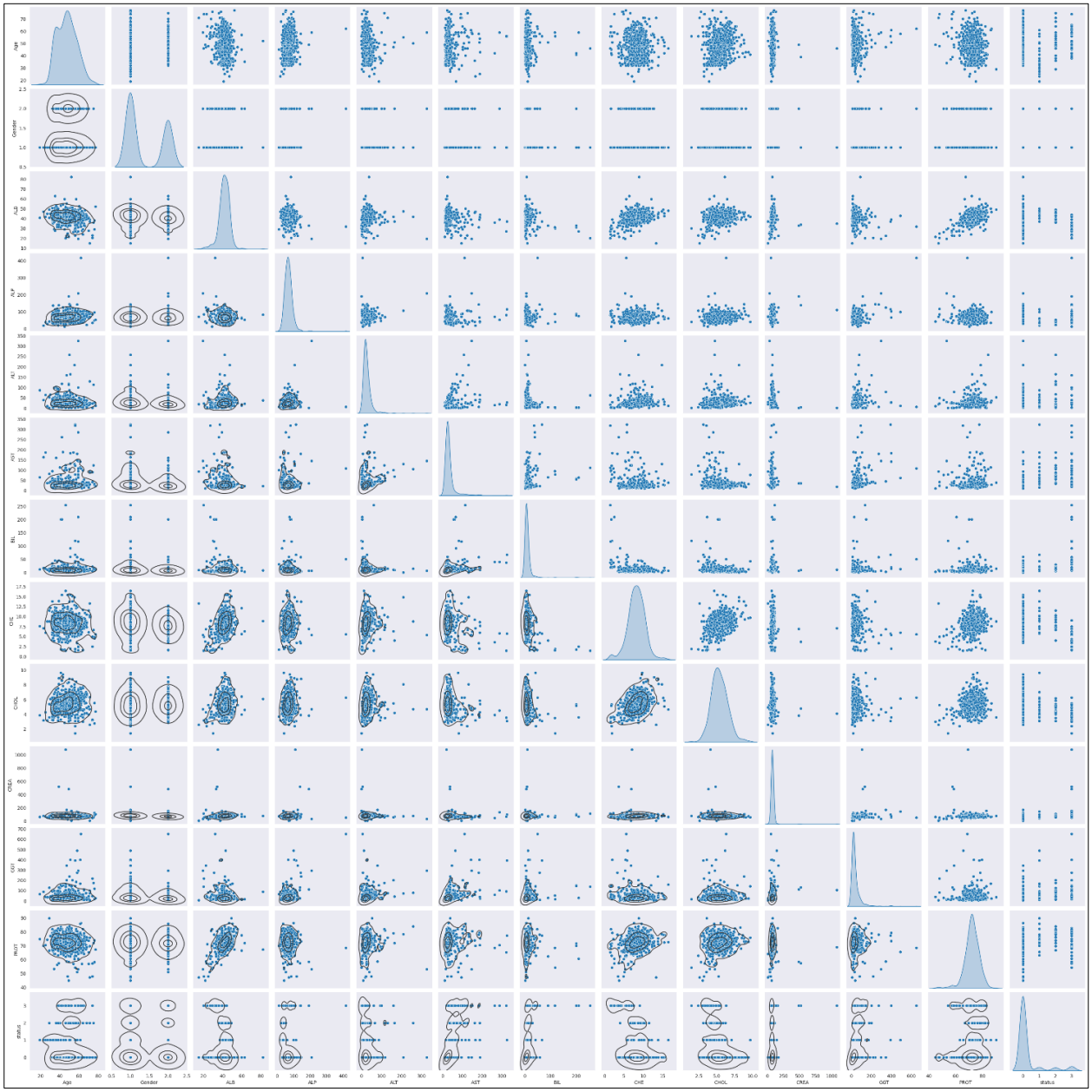


Figure 5. Pair plot of the features

Table 2. Parameters setting in this study

#	Machine Learning Algorithms	Parameter setting
1	AdaBoost	n_estimators=100, learning_rate=1.0
2	Random Forest	max_depth=8, max_leaf_nodes=11, n_estimators=390, min_samples_split=2
3	Rotation Forest	n_estimators=166, min_group=3, max_group=3
4	XGBoost	alpha=10, max_depth=12, learning_rate=1.0, n_estimators=40
5	LightGBM	colsample_bytree=0.8, max_depth=4, min_child_samples=5, n_estimators=1000

3.3.1. AdaBoost

AdaBoost, which stands for Adaptive Boosting, was introduced in 1995 and works by sequentially training a series of decision trees with limited depth on repeatedly modified versions of the dataset [29]. AdaBoost assigns higher weights to misclassified instances during each iteration, forcing subsequent models to focus more on the difficult-to-classify examples. At the end of the training, AdaBoost combines the predictions of all weak learners, giving more weight to models' predictions with higher accuracy [30]. The final ensemble prediction is determined by a weighted sum of the individual weak learner predictions, where models with higher accuracy contribute more to the final decision [31].

3.3.2. Random Forest

The Random Forest algorithm, introduced in 2001, is an ensemble learning method that constructs a multitude of

decision trees during training [32]. During the construction of each tree, a random subset of features is selected as candidates for splitting at each node, which leads to decorrelation of the individual trees and improves them within the ensemble [33].

3.3.3. Rotation Forest

The Rotation Forest algorithm was introduced in 2006 [34]. It is an ensemble learning method that combines Principal Component Analysis (PCA) principles and decision tree ensembles. During training, Rotation Forest first applies PCA to rotate the feature space, creating new feature subsets that capture different aspects of the data. Eventually, decision trees create a strong intergraded ensemble classifier.

3.3.4. XGBoost

XGBoost, which stands for Extreme Gradient Boosting, was introduced in 2016 [35]. XGBoost is a gradient-boosting algorithm that iteratively builds a series of decision trees to predict the target variable. During each iteration, XGBoost minimizes a specific loss function by fitting a new tree to the residual errors of the previous model [36].

3.3.5. LightGBM

The LightGBM or Light Gradient Boosting algorithm, which was introduced in 2017, is also a gradient boosting framework as XGBoost that uses a Gradient-based One-Side Sampling (GOSS) to reduce memory usage and improve training speed. During training, LightGBM partitions the dataset vertically rather than horizontally, allowing it to use only a subset of the data for calculating gradients. Using GOSS, the algorithm ensures that important data points are retained, and by employing histogram-based algorithms, LightGBM finds the best split points for enhancing efficiency and scalability [37].

3.4. Performance Metrics

To understand and compare the efficacy of algorithms, we'll need performance metrics to evaluate the performance and efficacy of these ensemble-based algorithms by using four performance metrics introduced in this section. These metrics include accuracy, precision, recall, and F1 score.

3.4.1. Accuracy

Accuracy is a metric used to evaluate the performance of a classification model. It measures the proportion of correctly classified instances out of all instances in the dataset.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

3.4.2. Precision

Precision is a metric used to evaluate the proportion of correctly predicted positive instances out of all instances predicted as positive by the model.

$$\text{Percision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

3.4.3.Recall

Recall is used to evaluate the performance of a classification model. It measures the proportion of correctly predicted positive instances out of all actual positive instances in the dataset.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

3.4.4. F1-score

The F1-score combines precision and recall into a single score, providing a balanced measure of the model's overall accuracy in the classification model.

$$\text{F1} - \text{score} = \frac{2\times\text{Percision}\times\text{Recall}}{\text{Percision+ Recall}} \quad (4)$$

4. Results

This research proposed that ensemble-based algorithms are more efficient than individual learners. To substantiate this claim, this study embarked on an empirical investigation employing five ensemble algorithms, including AdaBoost, Random Forest, Rotation Forest, XGBoost, and LightGBM. The objective of the study was to validate this hypothesis by assessing the performance of these algorithms across various metrics. The culmination of the findings is encapsulated in Table 3, where a comprehensive summary of the results are presented.

Upon scrutinizing the outcomes, it became evident that all algorithms under examination demonstrated commendable performance across key evaluation metrics, including accuracy, precision, recall, and F1-score. Notably, LightGBM emerged as the frontrunner, boasting an impressive accuracy, precision, recall, and F1-score of 98.37%. Precision is still an important metric because it provides insight into the reliability of positive predictions made by the model. A high precision score indicates that the model is making fewer false positives, revealing that LightGBM results are reliable. Following closely behind was Rotation Forest, securing the second position with a noteworthy accuracy of 96.74%. Further down the rankings, XGBoost clinched the third position with an accuracy of 95.12%, followed by Random Forest in fourth place with an accuracy of 94.19%. AdaBoost garnered an accuracy of 93.30%, and placed the fifth position. These results underscore the exceptional performance of LightGBM compared to its ensemble counterparts. Figure 6 represents algorithm performance in a bar plot.

Moreover, the current research extends beyond mere performance evaluation, as it endeavours to contextualize the findings within the broader landscape of existing studies. Table 4 presents a comparison of the results with those of other studies. Despite disparities in datasets and experimental setups, the analysis reveals a consistent trend: LightGBM consistently outperforms other algorithms across multiple evaluation metrics. This recurrent pattern underscores the robustness and efficacy of LightGBM

across diverse research contexts. Table 4 represents a comparison with other studies on the same dataset.

Table 3. Performance results of the ensemble-based algorithms used in this study

#	Machine Learning Algorithms	Accuracy Score (%)	Precision Score (%)	Recall Score (%)	F1 Score (%)
1	AdaBoost	93.30	93.30	93.30	93.30
2	Random Forest	94.19	94.19	94.19	94.19
3	Rotation Forest	96.74	96.74	96.74	96.74
4	XGBoost	95.12	95.12	95.12	95.12
5	LightGBM	98.37	98.37	98.37	98.37

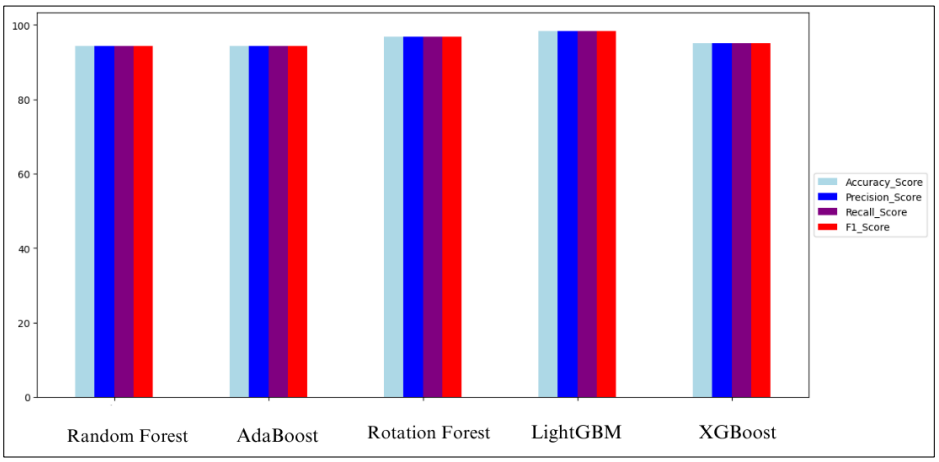


Figure 6. Algorithms performance bar plot

Table 4. Comparison with other studies on the same dataset

Reference	Machine Learning Algorithms	Accuracy Score (%)
This study	AdaBoost	93.30
	Random Forest	94.19
	Rotation Forest	96.74
	XGBoost	95.12
	LightGBM	**98.37**
[24]	K-nearest Neighbors (KNN)	89
	Decision Tree	92
	ANN	92
	SVM	95
	Logistic Regression	94
	XGBoost	95
[25]	Logistic regression	95.67
	Naive Bayes	92.43
	SVM	94.59
	K-nearest neighbors (K = 5)	95.67
	Decision tree	96.75
	Random forest	97.29
	Logistic regression	95.67

5. Conclusion

This study sheds light on the critical role of ensemble-based algorithms in the early detection and classification of liver diseases, including Hepatitis, Fibrosis, and Cirrhosis. The liver's significance as the largest internal organ in the human body cannot be overstated, as it orchestrates over 500 metabolic activities crucial for maintaining bodily functions. However, the emergence of the Hepatitis C Virus presents a formidable challenge to liver health, underscoring the urgent need for early identification of liver diseases to prevent their progression to carcinoma and

potentially save lives. After pre-processing and allocating 80% of the dataset for training, five ensemble-based algorithms—AdaBoost, Random Forest, Rotation Forest, XGBoost, and LightGBM—were trained and evaluated. These algorithms were assessed thoroughly across four key performance metrics, including accuracy, precision, recall, and F1-score. Remarkably, LightGBM emerged as the standout performer, boasting. The superior performance of LightGBM underscores its potential as a promising algorithm for detecting and classifying liver diseases. This research underscores the importance of leveraging advanced machine learning techniques, particularly ensemble-based algorithms, to enhance early detection, improve patient outcomes, and facilitate more effective management of liver-related ailments in clinical settings. The findings of this study highlight the importance of early identification in mitigating the adverse effects of liver diseases. Timely intervention enabled by accurate detection algorithms can prevent the progression of ailments, reducing the burden on healthcare systems and improving the quality of life for affected individuals

6. References

[1] Solomon, E. P., Berg, L. R., Martin, D. W. (2014). Biology. Brooks/Cole Thomson Learning, Boston, United States.

[2] Rehermann, B., & Nascimbeni, M. (2005). Immunology of hepatitis B virus and hepatitis C virus infection. Nature Reviews Immunology, 5(3), 215–229. doi:10.1038/nri1573.

[3] Institute for Quality and Efficiency in Health Care (IQWiG) (2023). In brief: How does the liver work?. Institute for Quality and Efficiency in Health Care (IQWiG), New York, United States. Available online: https://www.ncbi.nlm.nih.gov/books/NBK279393/ (accessed on August 2024).

[4] Park, S.-J., & Hahn, Y. S. (2023). Hepatocytes infected with hepatitis C virus change immunological features in the liver microenvironment. Clinical and Molecular Hepatology, 29(1), 65–76. doi:10.3350/cmh.2022.0032

[5] Campollo, O., Amaya, G., & McCormick, P. A. (2022). Milestones in the discovery of hepatitis C. World Journal of Gastroenterology, 28(37), 5395–5402. doi:10.3748/wjg.v28.i37.5395.

[6] [6] Ramos-Tovar, E., & Muriel, P. (2020). Molecular Mechanisms That Link Oxidative Stress, Inflammation, and Fibrosis in the Liver. Antioxidants, 9(12), 1279. doi:10.3390/antiox9121279.

[7] Mohd Hanafiah, K., Groeger, J., Flaxman, A. D., & Wiersma, S. T. (2013). Global epidemiology of hepatitis C virus infection: New estimates of age-specific antibody to HCV seroprevalence. Hepatology, 57(4), 1333–1342. doi:10.1002/hep.26141.

[8] Harabor, V., Mogos, R., Nechita, A., Adam, A.-M., Adam, G., Melinte-Popescu, A.-S., Melinte-Popescu, M., Stuparu-Cretu, M., Vasilache, I.-A., Mihalceanu, E., Carauleanu, A., Bivoleanu, A., & Harabor, A. (2023). Machine Learning Approaches for the Prediction of Hepatitis B and C Seropositivity. International Journal of Environmental Research and Public Health, 20(3), 2380. doi:10.3390/ijerph20032380.

[9] Zhang, C., Shu, Z., Chen, S., Peng, J., Zhao, Y., Dai, X., Li, J., Zou, X., Hu, J., & Huang, H. (2024). A machine learning-based model analysis for serum markers of liver fibrosis in chronic hepatitis B patients. Scientific Reports, 14(1), 12081. doi:10.1038/s41598-024-63095-8.

[10] Elshewey, A. M., Shams, M. Y., Tawfeek, S. M., Alharbi, A. H., Ibrahim, A., Abdelhamid, A. A., Eid, M. M., Khodadadi, N., Abualigah, L., Khafaga, D. S., & Tarek, Z. (2023). Optimizing HCV Disease Prediction in Egypt: The hyOPTGB Framework. Diagnostics, 13(22), 3439. doi:10.3390/diagnostics13223439.

[11] Feng, S., Wang, J., Wang, L., Qiu, Q., Chen, D., Su, H., Li, X., Xiao, Y., & Lin, C. (2023). Current Status And Analysis Of Machine Learning in Hepatocellular Carcinoma. Journal of Clinical and Translational Hepatology, 11(5), 1184–1191. doi:10.14218/JCTH.2022.00077S.

[12] Arjmand, A., Angelis, C. T., Tzallas, A. T., Tsipouras, M. G., Glavas, E., Forlano, R., Manousou, P., & Giannakeas, N. (2019). Deep Learning in Liver Biopsies using Convolutional Neural Networks. 2019 42nd International Conference on Telecommunications and Signal Processing (TSP). doi:10.1109/tsp.2019.8768837.

[13] Lu, M.-Y., Huang, C.-F., Hung, C.-H., Tai, C., Mo, L.-R., Kuo, H.-T., Tseng, K.-C., Lo, C.-C., Bair, M.-J., Wang, S.-J., Huang, J.-F., Yeh, M.-L., Chen, C.-T., Tsai, M.-C., Huang, C.-W., Lee, P.-L., Yang, T.-H., Huang, Y.-H., … Chong, L.-W. (2023). Artificial intelligence predicts direct-acting antivirals failure among hepatitis C virus patients: A nationwide hepatitis C virus registry program. Clinical and Molecular Hepatology, 30(1), 64–79. doi:10.3350/cmh.2023.0287.

[14] Park, H., Lo-Ciganic, W. H., Huang, J., Wu, Y., Henry, L., Peter, J., Sulkowski, M., & Nelson, D. R. (2022). Machine learning algorithms for predicting direct-acting antiviral treatment failure in chronic hepatitis C: An HCV-TARGET analysis. Hepatology, 76(2), 483–491. doi:10.1002/hep.32347.

[15] Akella, A., & Akella, S. (2020). Applying machine learning to evaluate for fibrosis in chronic hepatitis c. MedRxiv, 2020-11. doi:10.1101/2020.11.02.20224840.

[16] Maiellaro, P., Cozzolongo, R., & Marino, P. (2005). Artificial Neural Networks for the Prediction of Response to Interferon Plus Ribavirin Treatment in Patients with Chronic Hepatitis C. Current Pharmaceutical Design, 10(17), 2101–2109. doi:10.2174/1381612043384240.

[17] Edeh, M. O., Dalal, S., Dhaou, I. Ben, Agubosim, C. C., Umoke, C. C., Richard-Nnabu, N. E., & Dahiya, N. (2022). Artificial Intelligence-Based Ensemble Learning Model for Prediction of Hepatitis C Disease. Frontiers in Public Health, 10, 892371. doi:10.3389/fpubh.2022.892371.

[18] Prakash, N. N., Rajesh, V., Namakhwa, D. L., Dwarkanath Pande, S., & Ahammad, S. H. (2023). A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis. Scientific African, 20(e01629), 1629. doi:10.1016/j.sciaf.2023.e01629.

[19] Park, H., Lo-Ciganic, W. H., Huang, J., Wu, Y., Henry, L., Peter, J., Sulkowski, M., & Nelson, D. R. (2022). Evaluation of machine learning algorithms for predicting direct-acting antiviral treatment failure among patients with chronic hepatitis C infection. Scientific Reports, 12(1), 18094. doi:10.1038/s41598-022-22819-4.

[20] Wang, Y., Yin, B., & Zhu, Q. (2023). Application of Machine Learning Algorithms in Predicting Hepatitis C. Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science. doi:10.1145/3644116.3644176.

[21] Butt, M. B., Alfayad, M., Saqib, S., Khan, M. A., Ahmad, M., Khan, M. A., & Elmitwally, N. S. (2021). Diagnosing the Stage of Hepatitis C Using Machine Learning. Journal of Healthcare Engineering, 2021, 8062410. doi:10.1155/2021/8062410.

[22] Zhang, L., Wang, J., Chang, R., & Wang, W. (2024). Investigation of the effectiveness of a classification method based on improved DAE feature extraction for hepatitis C prediction. Scientific Reports, 14(1), 9143. doi:10.1038/s41598-024-59785-y.

[23] Lilhore, U. K., Manoharan, P., Sandhu, J. K., Simaiya, S., Dalal, S., Baqasah, A. M., Alsafyani, M., Alroobaea, R., Keshta, I., & Raahemifar, K. (2023). Hybrid model for precise hepatitis-C classification using improved random forest and SVM method. Scientific Reports, 13(1). doi:10.1038/s41598-023-36605-3.

[24] Alizargar, A., Chang, Y. L., & Tan, T. H. (2023). Performance Comparison of Machine Learning Approaches on Hepatitis C Prediction Employing Data Mining Techniques. Bioengineering, 10(4). doi:10.3390/bioengineering10040481.

[25] Safdari, R., Deghatipour, A., Gholamzadeh, M., & Maghooli, K. (2022). Applying data mining techniques to classify patients with suspected hepatitis C virus infection. Intelligent Medicine, 2(4), 193–198. doi:10.1016/j.imed.2021.12.003.

[26] Lichtinghagen, R., Klawonn, F., & Hoffmann, G. (2020). HCV data. UCI Machine Learning Repository, 10, C5D612.

[27] Hair, J. F., Babin, B. J., Black, W. C., Anderson, R. E. (2019). Multivariate Data Analysis. Cengage, Boston, United Kingdom.

[28] ickham, H., Grolemund, G. (2016). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media, Sebastopol, United States.

[29] Schapire, R.E. (2003). The Boosting Approach to Machine Learning: An Overview. Nonlinear Estimation and Classification. Lecture Notes in Statistics, 171, Springer, New York, United States. doi:10.1007/978-0-387-21579-2_9.

[30] Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). The Annals of Statistics, 28(2). doi:10.1214/aos/1016218223.

[31] Rätsch, G., Onoda, T., & Müller, K. R. (2001). Soft margins for AdaBoost. Machine Learning, 42(3), 287–320. doi:10.1023/A:1007618119488.

[32] Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. Ecology, 88(11), 2783–2792. doi:10.1890/07-0539.1.

[33] Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. Multiple Classifier Systems, MCS 2000, Lecture Notes in Computer Science, 1857, Springer, Berlin, Germany. doi:10.1007/3-540-45014-9_1.

[34] Rodríguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation forest: A New classifier ensemble method. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(10), 1619–1630. doi:10.1109/TPAMI.2006.211.

[35] Chen, T., & Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. doi:10.1145/2939672.2939785.

[36] Hastie, T., Friedman, J., & Tibshirani, R. (2001). The Elements of Statistical Learning. Springer Series in Statistics, Springer New York, United States. doi:10.1007/978-0-387-21606-5.

[37] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 4-9 December, 2017, Long Beach, United States.