Available online at www.medicinescience.org

ORIGINAL ARTICLE

Medicine Science International Medical Journal

Medicine Science 2025;14(4):1309-17

# Investigation of Hepatitis C diagnosis with machine learning and evaluation of clinical biomarkers with explainable artificial intelligence models

**Fatma Hilal Yagin[1], Abdulvahap Pinar[2]**

*[1]İnönü University, Faculty of Medicine, Biostatistics and Medical Informatics, Malatya, Türkiye*
*[2]Adıyaman University, Rectorate Unit, Adıyaman, Türkiye*

Received 14 April 2025; Accepted 02 July 2025
Available online 30 November 2025 with doi: 10.5455/medscience.2025.04.092



**Abstract**

This study aimed to evaluate the performance of machine learning (ML) models in the diagnosis of Hepatitis C Virus (HCV) patients and to identify clinical biomarkers using explainable artificial intelligence (XAI) approaches. Black box algorithms - Random Forest (RF) and Extreme Gradient Boosting (XGBoost) were used in the study, and XAI methods - SHapley Additive Explanations (SHAP) and Accumulated Local Effects (ALE) were applied to increase the interpretability of the models. There were 615 patients in the dataset, and the output variable included various liver disorders including HCV. The results showed that RF and XGBoost models exhibited high performance with 93.75% and 92.38% accuracy rates, respectively. SHAP and ALE analyses revealed the importance and interactions of the factors (ALT, AST, bilirubin, albumin, age) underlying model decisions. This study demonstrates the potential of ML models in early diagnosis of HCV infection and how they can be integrated with XAI methods to make them more reliable in medical applications.

**Keywords:** Hepatitis C diagnosis, machine learning, explainable artificial intelligence, shapley additive explanations, accumulated local effects

## Introduction

Hepatitis C Virus (HCV) is recognized as a major public health problem worldwide and is one of the most common causes of liver disease. Long-term complications include life-threatening conditions such as cirrhosis and hepatocellular carcinoma. Early diagnosis of HCV infection is crucial to improve treatment success and prevent progressive liver damage. However, the limitations of traditional diagnostic methods and the complex underlying biological processes make it difficult to accurately identify HCV infection [1-4]. HCV affects approximately 71 million people worldwide, with the highest prevalence in the Eastern Mediterranean and European regions [5]. Transmission is primarily through exposure to infected blood, with risk factors including injection drug use (accounting for the majority of new infections in high-income countries), inadequate infection control during healthcare procedures, unscreened blood transfusions (especially in low-income countries), and, less commonly, sexual transmission and mother-to-child transmission [6]. The infection has a remarkable chronicity rate of 55–85%, with geographic variation in genotype distribution—genotype 1 predominates in North America and Europe, while genotypes 3 and 4 are more common in South Asia and Africa, respectively. Despite the availability of highly effective direct-acting antiviral therapies that provide cure rates exceeding 95%, significant public health challenges remain, including late diagnosis due to the asymptomatic nature of early infection, limited access to testing and treatment in many regions, and the lack of an effective vaccine. These factors contribute to HCV remaining a major cause of liver cirrhosis and hepatocellular carcinoma globally [5].

Machine Learning (ML) is increasingly used in the development of decision support systems in healthcare. In particular, for the diagnosis of complex diseases such as HCV, ML models

**CITATION**
Yagin FH, Pinar A. Investigation of Hepatitis C Diagnosis with Machine Learning and Evaluation of Clinical Biomarkers with Explainable Artificial Intelligence Models. Med Science. 2025;14(4):1309-17.

**Corresponding Author:** Fatma Hilal Yagin, İnönü University, Faculty of Medicine, Department of Biostatics and Medical Informatics, Malatya, Türkiye
Email: hilal.yagin@inonu.edu.tr

DOI: 10.5455/medscience.2025.04.092

can provide highly accurate diagnostic models by utilizing large data sets. However, the interpretability of the decision processes of these so-called "black box" models is critical for reliability in medical applications. Clinicians want to understand how ML models that influence treatment decisions work, as this is essential to avoid misdiagnosis and maintain patient trust. Therefore, XAI methods are used to explain the complex decision processes of ML models and generate meaningful interpretations. XAI is a set of methods developed to make the decision processes of ML models transparent and understandable by humans. It reveals how traditional "black box" models generate predictions, which data features play a critical role in decisions, and how these features interact with each other [7-12].

In this study, Explainable Artificial Intelligence (XAI) techniques such as SHapley Additive Explanations (SHAP) and Accumulated Local Effects (ALE) are applied to reveal how Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models make decisions. SHAP shows the degree of contribution of each feature to model predictions and the interactions between features, while ALE analyzes the local effects of individual features on the target variable and decomposes the "black box" structure of the model. With these approaches, medical experts have the tools to understand which biological parameters the model prioritizes and how they affect disease risk [13-15].

The aim of this study is to evaluate the performance of ML models in diagnosing liver diseases in the HCV dataset and to integrate XAI methods to explain the decision processes of these models. Within the scope of the study, the performance of the models built using RF and XGBoost algorithms were compared and the interpretability of the models was improved with SHAP and ALE analyses. The potential of ML-based approaches in the diagnosis of HCV infection and recommendations for their reliable integration in medical applications are presented.

## Material and Methods

### Dataset and related factors

The purpose of this study is to look at the feasibility of predicting the characteristics and risk factors of HCV. The output variable (diagnosis) is classified into the following categories: Blood Donor (Healthy persons) 533 (86.7%), Suspect Blood Donor (persons with possible liver disease) 7 (1.1%), Hepatitis 24 (3.9%), Fibrosis 21 (3.4%), and Cirrhosis 30 (4.9%), for a total of 615 samples. The open-access dataset utilized in the research was acquired from the following source: https://archive.ics.uci.edu/dataset/571/hcv+data [16]. The variable's include Gender, Cholesterol (CHOL), Creatinine (CREA), Age, Alanine aminotransferase (ALT), Gamma-glutamyl transferase (GGT), Alkaline phosphatase (ALP), Total protein (PROT), Cholinesterase (CHE), Bilirubin (BIL), Albumin (ALB), and Aspartate aminotransferase (AST). As this study was based exclusively on publicly available open-access datasets, ethical committee approval was not required. RF and XGBoost black box algorithms from ML models were

**Table 1.** Input and output characteristics of Hepatitis C Virus

| Variable/Feature | Description | Type | Role |
|---|---|---|---|
| **Diagnosis** | Health status of the individual (0=Blood Donor,1=suspect Blood Donor, 2=Hepatitis, 3=Fibrosis, 4=Cirrhosis) | Categorical | Output |
| **Age** | Individual's age | Quantitative | Input |
| **Sex** | 1=Male, 2=Female | Categorical | Input |
| **ALB** | Albumin | Quantitative | Input |
| **ALP** | Alkaline phosphatase | Quantitative | Input |
| **ALT** | Alanine aminotransferase | Quantitative | Input |
| **AST** | Aspartate aminotransferase | Quantitative | Input |
| **BIL** | Bilirubin | Quantitative | Input |
| **CHE** | Cholinesterase | Quantitative | Input |
| **CHOL** | Cholesterol | Quantitative | Input |
| **CREA** | Creatinine | Quantitative | Input |
| **GGT** | Gamma-glutamyl transferase | Quantitative | Input |
| **PROT** | Total protein | Quantitative | Input |

DOI: 10.5455/medscience.2025.04.092

employed with these variables, and SHAP, and ALE plots were constructed to assess the model's interpretability, as well as the contributions and interactions of the variables to model predictions. Table 1 displays the variables' inputs and outputs.

**Biostatistical data analysis**

Median and interquartile range were used to summarize quantitative variables. For variables not normally distributed, the Kruskal-Wallis H test was used to investigate associations between quantitative factors and HCV diagnosis. Bonferroni correction was used for multiple comparisons between groups with significant differences. Pearson Chi-Square test was used to examine the relationship between qualitative variables. The model performances of Random forest and XGBoost algorithms from machine learning models were calculated and the model was interpreted with explainable artificial intelligence methods. In all hypothesis test results, $p < 0.05$ was considered statistically significant. Statistical analyses were performed with SPSS software version 27.0. Python version 3.12.6 was used for machine learning analysis and associated tables.

**Machine Learning Methodologies**

ML is a programming approach inspired by AI that allows computers to build and improve their learning capacities via data intake [17]. Arthur Samuel, who became famous for his work on the game of checkers, coined the phrase "ML" in 1959 at IBM. He characterized ML as a vast scientific area of research in which computers may approach a problem in the same way that humans would, without having to be explicitly taught each time. ML is used to educate computers how to utilize data effectively and constructively, making the end findings simpler to understand. It is a collection of algorithms and statistical models that, when applied on computer systems, can do specified jobs automatically [18]. In this study, classification models were created using RF and XGBoost algorithms to distinguish HCV patients. A holdout approach with 100 separate random iterations was used to test the predictive models. The holdout approach is a methodology to determine the generalizability of an ML model to previously unknown data. It works by splitting the dataset into training and test sets, then training and testing the model on a portion of the data. The dataset was first split into training and test sets in a 4-1 ratio, and this process was repeated 100 times. Finally, the validation procedure was performed by splitting the dataset into 100 training and test sets using 100 different random seeds, and the model was trained and tested on each seed. The average values for these 100 iterations were calculated to produce a reliable prediction [19, 20].

RF and XGBoost were chosen as classification methods, while SHAP and ALE were selected as explainability methods, due to their mutually beneficial strengths in achieving the goals of the study. RF was utilized for the reasons of its strong ensemble learning system that combats overfitting through bagging and random feature selection, its effectiveness in handling missing data, and its ability to identify important biomarkers. The chosen algorithm for the best possible predictive performance was XGBoost, a gradient-boosted decision tree algorithm, which is capable of optimizing a sequence of weak learners, able to handle missing values, with high predictive accuracy on arbitrary datasets. To enable interpretability, we used SHAP in a model-agnostic manner, whereby the contribution of each clinical variable (e.g., ALT, AST) to HCV diagnosis draws from cooperative game theory principles to ensure transparency in model predictions. ALE, in contrast, provided concise, interaction-agnostic global explanations, showing how specific ranges of biomarkers had a direct influence on the models predictions. Collectively, these algorithms offered a balance between predictive power and explainability, matching the aim of the study to develop reliable interpretable ML models for clinical settings.

***Random Forest***

RF is a popular ensemble learning method for regression and classification in ML. It increases the precision and reliability of predictions, improves the generalizability of the model, and reduces the risk of overfitting by training each tree with a randomly selected set of data points. RF is the solution to deal with missing data, find important features, and avoid outliers. Due to these features, it finds widespread application in evaluating large data sets and making predictions in various fields [21]. This method dramatically eliminates overfitting, a major issue with individual decision trees, while increasing overall prediction accuracy. This resilience is critical for trustworthy outcomes in medical datasets like HCV, which may include nuanced patterns.

***Extreme Gradient Boosting***

XGBoost is a decision tree-based ensemble approximation method. It sequentially trains multiple weak tree models to produce a robust learner by maximizing an objective function at each step. The XGBoost model builds a prediction model by combining multiple weak learners. Each model derives knowledge from its predecessor and builds a robust model by sequentially changing the weights. XGBoost provides the potential to automatically handle missing data which can be valuable for infrastructure health monitoring when large missing values are present [22, 23]. This work used XGBoost to construct a high-performance classification model using the HCV dataset and to elucidate the model's decision-making processes.

**XAI Methodologies**

ML has achieved extensive acceptance and has been used across several fields and applications. Nevertheless, certain procedures must be enacted to guarantee societal acceptance and confidence in ML powered systems. To establish this confidence, it is essential to elucidate and demonstrate the decision-making processes of ML models. XAI facilitates user comprehension of an algorithm's decision-making process and the training data used [10]. This research employs two explainable artificial intelligence

DOI: 10.5455/medscience.2025.04.092

approaches, SHAP and ALE. SHAP analysis elucidated the contribution of each trait to the model's predictions, highlighting trait significance and interdependencies, while ALE analysis scrutinized the influence of characteristics on the diagnostic variable with more specificity. These techniques elucidated the model's "black box" structure, enabling us to understand and analyze the elements influencing the model's decision-making process. Consequently, while enhancing the model's trustworthiness, it also furnished professionals with critical insights into its functionality [24-26].

### *Shapley Additive Explanations*

We utilized ML-based feature importance approaches to determine the value of features and explain how the model predicts complicated models. We wish to discover which risk variables (features) have the strongest relationship with the result (diagnosis). Each input characteristic is assigned a significance value for a specific prediction using SHAP, a model-independent explanation. The foundation of SHAP is the important value derived by probabilistically estimating each player's contribution to the game's ultimate result using the Shapley value and the concepts of cooperative game theory [27].

### *Accumulated Local Effects*

ALE was first established as a model-independent strategy for global explanations of black-box ML algorithm results [28]. ALE offers a significant advantage over other techniques such as partial dependence plots and SHAP: its results provide a clear functional decomposition of the model [29]. As a result, the presence or absence of interactions between variables in a mode has no impact on ALE results. Another plus is how fast it can be calculated [30].

## Results

Statistical information on quantitative variables is shown in Table 2. Median values for variables are summarized with interquartile ranges (IQR). In the HCV dataset, there is a significant difference ($p<0.001$) in the means of quantitative variables (Age, ALB:Albumin, ALP: Alkaline phosphatase, ALT: Alanine aminotransferase, AST: Aspartate aminotransferase, BİL: Bilirubin, CHE: Cholinesterase, CHOL: Cholesterol, CREA: Creatinine, GGT: Gamma-glutamyltransferase, PROT: Total protein) between diagnostic groups (Blood Donor, Suspected Blood Donor, Hepatitis, Fibrosis, Cirrhosis, Blood Donor).

Table 3 shows descriptors for the diagnosis group according to gender. The percentage of women (59.70%) in the Blood Donor category was somewhat greater than that of males. The percentage of women in the Suspect Blood Donor and Hepatitis categories, however, was much higher (85.70% and 83.30%, respectively). In both the Fibrosis and Cirrhosis categories, women outnumbered males. The statistical study revealed no significant link between Gender (female and male) and Diagnosis categories ($p>0.05$; $p=0.106$).

**Table 2.** Descriptive statistics for quantitative variables

| Variables* | Diagnosis | | | | | p |
|---|---|---|---|---|---|---|
| * | Blood Donor | Suspect Blood Donor | Hepatitis | Fibrosis | Cirrhosis | |
| **Age** | 47$^{abcd}$ (14.5) | 55$^{b}$ (23) | 37$^{cd}$ (18.5) | 51 (11) | 56 (13) | <0.001*** |
| **ALB** | 42.2$^{ad}$ (6.2) | 21.6$^{bc}$ (5.6) | 43.5$^{d}$ (5.5) | 41d (6.5) | 33 (7.75) | <0.001*** |
| **ALP** | 66.7$^{abc}$ (24.55) | 106$^{bc}$ (54.2) | 36.95$^{d}$ (27.125) | 45.2$^{d}$ (28.7) | 67.9 (51.925) | <0.001*** |
| **ALT** | 23.1$^{d}$ (15.15) | 49.2$^{bd}$ (189.1) | 18.3 (29.6) | 34d (104.45) | 5.65 (24.225) | <0.001*** |
| **AST** | 24.8$^{abcd}$ (8.7) | 46.7 (106.3) | 47.2 (53.675) | 70 (67.8) | 92.9 (69.125) | <0.001*** |
| **BIL** | 6.9$^{abcd}$ (5.1) | 4.9$^{bcd}$ (4.6) | 13$^{d}$ (9.5) | 13 (5.5) | 34 (53.5) | <0.001*** |
| **CHE** | 8.35$^{d}$ (2.525) | 5.33$^{bd}$ (9.55) | 9.51$^{d}$ (3.008) | 8.59$^{d}$ (2.355) | 3.425 (4.11) | <0.001*** |
| **CHOL** | 5.38$^{acd}$ (1.445) | 4.3 (2.24) | 5.06$^{d}$ (1.845) | 4.64 (0.9) | 3.925 (1.365) | <0.001*** |
| **CREA** | 78$^{ab}$ (20) | 52 (47) | 72.25 (20.35) | 71.4 (14.35) | 68.5 (48.35) | 0.00325*** |
| **GGT** | 21.4$^{abcd}$ (16.95) | 83 (256.7) | 45.55 (58.25) | 72.2 (52.1) | 96.35 (94.475) | <0.001*** |
| **PROT** | 72.2$^{ac}$ (5.85) | 47.8$^{bcd}$ (11.6) | 73.65$^{d}$ (6.275) | 76.1$^{d}$ (8.8) | 70.25 (12.6) | <0.001*** |

*: a: Suspect Blood Donor is different according to group, b: Hepatitis is different according to group, c: Fibrosis is different according to the group, d: Cirrhosis is different according to group.
**: Variables, ' summarized as median; ***: Kruskal Wallis H test

DOI: 10.5455/medscience.2025.04.092

**Table 3.** Descriptive statistics for quantitative variables

| Variables | Categories | Gender | | P-value |
|---|---|---|---|---|
| | | Female N(%) | Male N(%) | |
| Diagnosis | **Blood Donor** | 318 (59.70%) | 215 (40.30%) | 0.106* |
| | **Suspect Blood Donor** | 6 (85.70%) | 1 (14.30%) | |
| | **Hepatitis** | 20 (83.30%) | 4 (16.70%) | |
| | **Fibrosis** | 13 (61.90%) | 8 (38.10%) | |
| | **Cirrhosis** | 20 (66.70%) | 10 (33.30%) | |

*: Pearson-Chi-Squired

The classification performance of the ML models used in the study is presented in Table 4. The RF and XGBoost models achieved an accuracy of 93.75% and 92.38%, respectively, indicating that HCV virus can be successfully detected in general. In addition, F1 scores (RF: 92.59% and XGBoost: 92.26%) reveal that the model exhibits a balanced performance in terms of both sensitivity and specificity. Area Under the Curve(AUC) values (RF: 98.6% and XGBoost: 98.07%) emphasize that the models can distinguish positive and negative samples well even with different threshold settings. While the sensitivity metrics (RF: 93.75% and XGBoost: 92.38%) indicate that a small number of HCV positive cases are missed, the low specificity rates (RF: 63.44% and XGBoost: 60.15%) indicate that the models falsely classify some negative samples as positive. RF slightly outperformed XGBoost in terms of accuracy, F1-score, AUC, and sensitivity, but both models exhibited similar weaknesses in specificity. These results highlight the limitations of "black box" models such as RF and XGBoost, despite their general success in detecting HCV, in terms of false-positive predictions. While high sensitivity provides an advantage in preventing virus jumps, low specificity highlights the need for additional validation steps in clinical applications.

**Table 4.** Evaluating the classification performance of two different ML models for HCV virus using various criteria

| Model | Accuracy | F1 Score | AUC | Sensitivity | Specificity |
|---|---|---|---|---|---|
| **Random Forest** | 93.75 | 92.59 | 98.6 | 93.75 | 63.44 |
| **XGBoost** | 92.38 | 92.26 | 98.07 | 92.38 | 60.15 |

*: Pearson-Chi-Squired

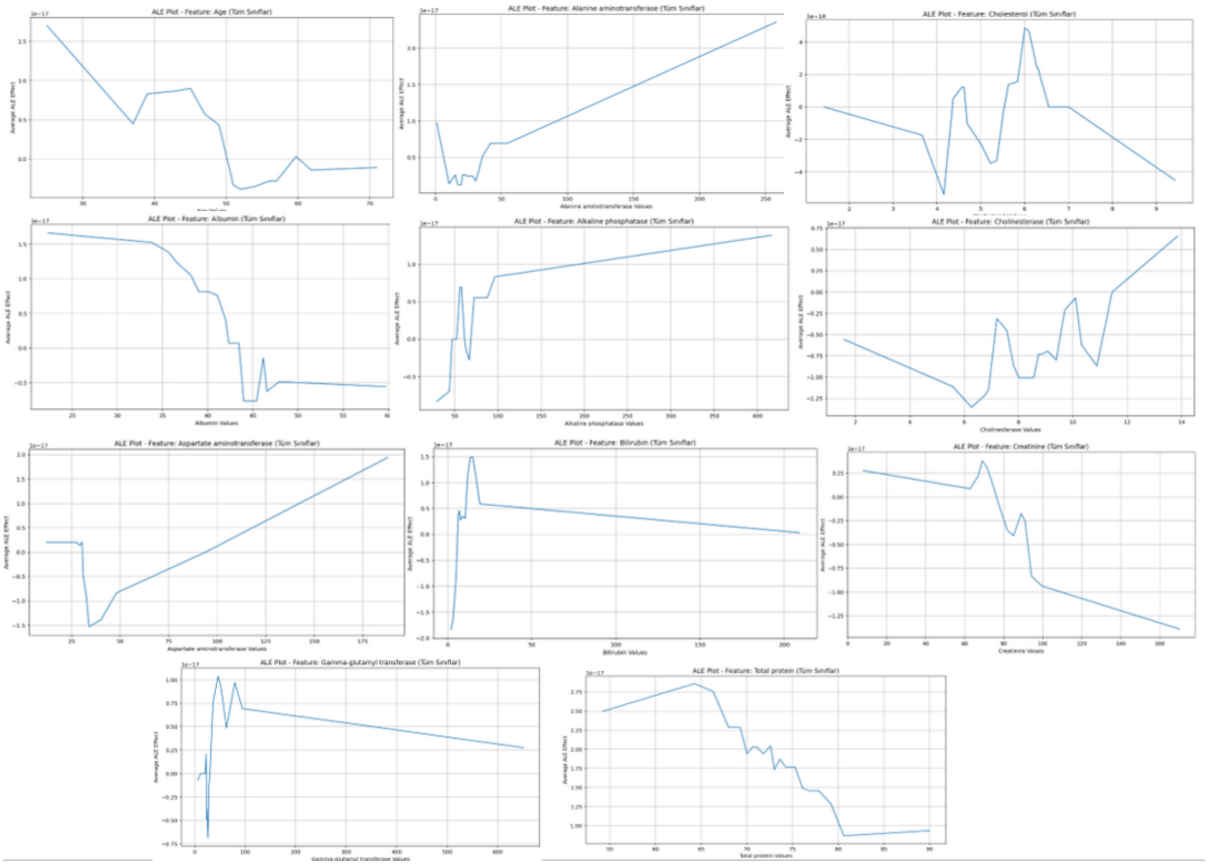


**Figure 1.** Examining the influence of factors on model predictions using ALE graphs

DOI: 10.5455/medscience.2025.04.092

When the ALE graphs in Figure 1 are examined, it is seen that the model prioritizes factors such as liver function tests (ALT, AST, Bilirubin, Fibrosis, Albumin and Age) in HCV diagnostic decisions. These variables, which represent liver damage, dysfunction and fibrosis, play an important role in the model's diagnostic choices. For example, reaching a certain threshold (i.e. the upper limit of the normal range) in ALT and AST levels increases the likelihood of the model predicting HCV positivity. This suggests that the model associate's liver damage with HCV infection and places a high value on these tests. Similarly, we found that low Albumin levels (i.e. the lower end of the normal range) increased the likelihood of the model making a positive prediction. As albumin is an important marker for liver dysfunction and fibrosis, the model's link between HCV infection and liver damage and fibrosis is supported. Furthermore, age appears to have a significant impact on the model's predictions within a certain range. This suggests that the model believes that age may be linked to HCV risk. These thresholds are very important in the model's decision-making procedures. This is because they predict that if these parameters are exceeded or reduced, the chances of the model producing a positive HCV prediction increase significantly.

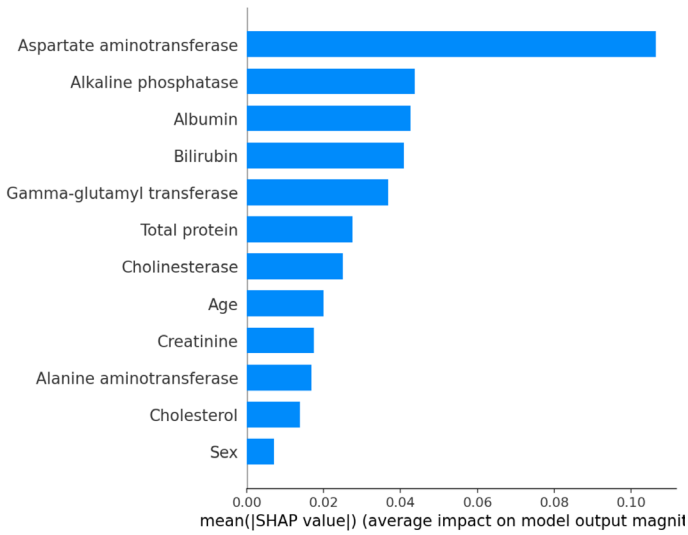


**Figure 2.** The magnitude of the impact of the SHAP mean plot on model outputs

Figure 2 shows the SHAP bar chart generated using the mean absolute SHAP value. Here, when all performance metrics were examined, the RF model was selected as the optimal prediction model and the optimal model descriptions were analyzed using the RF model. The SHAP chart helped us understand how the RF "black box" model works. According to the SHAP chart, it was determined that liver function tests such as AST, ALP, ALB, BIL, and GGT made the most significant contributions to the HCV prediction performance of the model. These results indicate that the model associates HCV infection with liver damage and function and is sensitive to these factors. The high accuracy and AUC values in Table 4 further confirm the strong performance of the model in general, while the SHAP chart shows which factors are more relevant in the decision-making processes of the model.

## Discussion

HCV infection is becoming a major worldwide health concern. HCV is widespread in many nations, posing a growing burden on society and healthcare systems. Long-term complications including cirrhosis and hepatocellular carcinoma are a significant issue [31]. Treatment providers for hepatitis HCV infections are realistically hoping that new therapeutic approaches will be more successful, more tolerated, and shorter in length than existing treatments due to the fast development of direct-acting antiviral medicines. [32,33]. Therefore, there is a need for rapid diagnostic tools for HCV.

In this study, we evaluated the performance of ML models (RF and XGBoost) in the diagnosis of HCV infection and used the XAI approaches SHAP and ALE to interpret their decision processes. The results showed that both RF and XGBoost models performed effectively in HCV diagnosis with high accuracy (RF: 93.75% and XGBoost: 92.38%), F1 score (RF: 92.59% and XGBoost: 92.26%) and AUC (RF: 98.6% and XGBoost: 98.07%). In particular, the high sensitivity values (RF: 93.75% and XGBoost: 92.38%) indicated that the models were successful in correctly identifying HCV positive cases, but the relatively low specificity rates (RF: 63.44% and XGBoost: 60.15%) revealed the risk of false positive diagnosis. This finding emphasizes the clinical criticality of preventing "false negatives" in serious health problems such as HCV. When comprehensive performance metrics were evaluated, the RF model performed better compared to XGBoost.

Analyses with XAI methods revealed that model decisions were based on biologically meaningful factors. SHAP and ALE plots showed that liver function tests such as ALT, AST, bilirubin, albumin, and age played a decisive role in model predictions. The most important contribution of this study is to make the decision processes of "black box" ML models interpretable for clinical experts. SHAP analysis quantified the contribution of each feature to the model, allowing clinicians to see which parameters are prioritized in the diagnosis. ALE plots guided the setting of clinical thresholds by showing how a given value (e.g. ALT $\geq$ 40 U/L) affected the model output. The explainability is critical for ML models to be used reliably in medical decisions.

HCV significantly affects liver function as evidenced by changes in liver enzymes and bilirubin levels. Basic liver function tests, including ALT, AST, bilirubin and albumin, serve as critical indicators of liver health in HCV patients. Elevated ALT and AST levels have been reported in HCV patients in the literature, with 85% of patients in the relevant study showing elevated ALT and 100% showing elevated AST. Total bilirubin was also elevated in 35% of HCV patients [34]. Another study applied treatment regimens such as interferon and ribavirin in HCV patients and reported significant reductions in ALT levels and improved liver function after treatment [35].

DOI: 10.5455/medscience.2025.04.092

The albumin-bilirubin (ALBI) score has been proposed as a useful indicator of liver function improvement after treatment [36]. Although elevated liver enzymes and bilirubin levels are indicative of liver damage in HCV, normalization of these parameters after treatment suggests the potential for improvement. However, some patients may still experience persistent liver dysfunction despite achieving sustained viral response, highlighting the complexity of HCV management [37]. When the literature was reviewed, the results of our SHAP and ALE-based XAI methodology were consistent with the literature.

This study shows that ML models, along with XAI methods such as SHAP and ALE, can provide a meaningful and accessible framework for diagnosing HCV infection. These models showed high accuracy (93.75% for RF; 92.38% for XGBoost) and sensitivity (>92%), indicating that they may have the potential to reduce false negative diagnoses which is particularly important for a virus associated with serious long-term complications such as cirrhosis and hepatocellular carcinoma. But the low specificity (60–63%) highlights the need to further refine model calibration to reduce false-positive classification, which can cause unnecessary anxiety, invasive testing, or overtreatment.

**To mitigate this limitation, future work should:**

Threshold Validation: Validation of model-derived thresholds (e.g., ALT ≥40 U/L, albumin ≤35 g/L) to established clinical guidelines to ensure that the diagnostic performance of thresholds, are not too unique to this population but rather aligned with real-world clinical thresholds already used in practice Working in conjunction with hepatologists would improve these cutoffs for various patient subgroups (ie, cohorts stratified according to age or comorbid conditions).

Multi-organ Biomarker Integration: Widening the data set to include indicators of renal function (eg, creatinine), inflammation (eg, CRP), and metabolic health (eg, HbA1c) may improve specificity. For example, incorporating hepatorenal indices (e.g., albumin-creatinine ratios) may capture systemic consequences of advanced HCV more appropriately.

Prospective Multicenter Validation: Furthering these analyses from a retrospective UCI dataset to a prospective, multicenter trials would further enhance generalizability. This measurement is essential to adjust for variation in HCV epidemiology according to demographic, genetic, and regional differences.

Insights gleaned from the study's XAI-driven modeling approach—prioritization of AST, ALP and bilirubin—are consistent with known clinical aspects of HCV pathophysiology, supporting biological validity of the models. These results have the potential to serve as a basis for developing automated decision-support tools for clinicians in providing rapid interpretation of liver function tests, thus limiting diagnostic delays. For instance, integrating ALE-based thresholds into electronic health records (EHRs) will facilitate flagging of patients at risk of developing disease for confirmatory RNA measurement.

Furthermore, the identification of age as a stray predictive factor (>50 year) is in concordance with the known data on immune senescence and delayed clearance of the virus, thus proposing age-stratified strategies to favor early detection. Although integrating omics data (e.g., genomic, proteomic) is a new scope for future research, this work lays the groundwork for interpretable ML for hepatology. Tools such as these enable the bridge from "black box" predictions to clinical decision making, likely to expedite the integration of precision medicine into HCV management, and reducing the global burden of liver disease.

## Limitations

The study has some limitations. First, the dataset was obtained from a single source (UCI) and analyzed retrospectively. The results need to be validated with prospective studies and larger patient populations. Second, the low specificity of the model requires careful interpretation of false positive results in clinical practice. In the future, the performance and interpretability of the model can be improved by adding more clinical parameters (genetic factors, lifestyle information) to the dataset. Furthermore, integrating omics data will enable significant advances in HCV diagnosis, monitoring and treatment for more comprehensive biomarker discovery.

## Conclusion

The study clearly shows the great potential of ML models in the diagnosis of the HCV. ML approaches have the potential to overcome the limitations of traditional diagnostic approaches, providing more rapid and potentially more accurate results through their ability to analyze large and complex datasets. Here, timely intervention and treatment initiation could give a significant advantage, especially in asymptomatic disease stages. ML different models like classification algorithms can be trained up to predict HCV patients that are positive and negative with high accuracy. This can enable healthcare professionals to conduct faster and more effective screenings helping to control HCV spread.

One more description of the study is that XAI contributes to the clinical applications of these ML models. The black box nature of AI systems' decision-making processes remains a significant barrier to the wider adoption of such systems in healthcare. SHAP, and ALE are examples of XAI techniques that help solve this problem by explaining how and why high-performance ML models make some decisions. SHAP explains the contribution of each feature to the model's output at the level of individual predictions, while ALE displays the average effect of a feature for an output. In this way, clinicians can transparently see and trust

the factors upon which the model settled to reach a diagnosis. These clear answers for the "why" and "how" questions also enable physicians to better assess the model's recommendations, identify potential errors, and biases, and ultimately make better-informed patient management decisions. These findings not only provide important details for HCV diagnosis but could also serve as a promising tool to be used for early diagnosis and personalized-treatment planning for other liver diseases. A timely and correct diagnosis in the initial phase of liver cancer or other chronic liver diseases can enhance the chance of its response to treatments. However, the proper integration of ML and XAI can lead to innovations in future management of such diseases.

**Conflict of Interests**
*The authors declare that there is no conflict of interest in the study.*

**Financial Disclosure**
*The authors declare that they have received no financial support for the study.*

**Ethical Approval**
*Due to the public availability of the data, ethical committee approval was unnecessary for this study.*

## References

1. Papatheodoridis G, Hatzakis A. Public health issues of hepatitis C virus infection. Best Pract Res Clin Gastroenterol. 2012;26:371-80.
2. Ghany MG, Morgan TR, AASLD-IDSA HCV Guidance Panel. Hepatitis C guidance 2019 update: AASLD–IDSA recommendations for testing, managing, and treating hepatitis C virus infection. Hepatology. 2020;71:686-721.
3. Harris R, Martin N, Rand E, et al. New treatments for hepatitis C virus: scope for preventing liver disease and transmission in England. J Viral Hepat. 2016;23:631-43.
4. Ansaldi F, Orsi A, Sticchi L, et al. Hepatitis C virus in the new era: perspectives in epidemiology, prevention, diagnostics, and predictors of therapy response. World J Gastroenterol. 2014;20:9633-57.
5. World Health Organization. Hepatitis B factsheet. Available from: www.who.int/news-room/fact-sheets/detail/hepatitis-b
6. Asselah T, Marcellin P, Schinazi RF. Treatment of hepatitis C virus infection with direct-acting antiviral agents: 100% cure? Liver Int. 2018;38:7-13.
7. Yağanoğlu M. Hepatitis C virus data analysis and prediction using machine learning. Data Knowl Eng. 2022;142:102087.
8. Bai FJ, Jasmine RA. Optimization of tree-based machine learning algorithms for improving the predictive accuracy of hepatitis C disease. In: Decision-Making Models. Elsevier; 2024. p.523-45.
9. Peng J, Zou K, Zhou M, et al. An explainable artificial intelligence framework for deterioration risk prediction in hepatitis patients. J Med Syst. 2021;45:61.
10. Alotaibi A, Alnajrani L, Alsheikh N, et al. Explainable ensemble-based machine learning models for detecting cirrhosis in hepatitis C patients. Computation. 2023;11:104.
11. Lu M-Y, Huang C-F, Hung C-H, et al. Artificial intelligence predicts direct-acting antiviral failure among hepatitis C virus patients: a nationwide registry program. Clin Mol Hepatol. 2023;30:64-75.
12. Sharma A, Khade T, Satapathy SM. A cross-dataset meta-model for hepatitis C detection using multi-dimensional pre-clustering. Sci Rep. 2025;15:7278.
13. Yang W, Fei J, Li J, et al. Environmental determinants of dynamic jogging patterns: insights from trajectory data and interpretable machine learning. Appl Geogr. 2025;178:103596.
14. Alabi RO, Elmusrati M, Leivo I, et al. Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP. Sci Rep. 2023;13:8984.
15. Shaon MSH, Karim T, Shakil MS, Hasan MZ. Comparative study of machine learning models with LASSO and SHAP feature selection for breast cancer prediction. Healthc Anal. 2024;6:100353.
16. UCI Machine Learning Repository. Available from: https://archive.ics.uci.edu/ml/datasets/Bank+Marketing 2017.
17. Mahesh B. Machine learning algorithms: a review. Int J Sci Res. 2020;9:381-6.
18. Drogkoula M, Kokkinos K, Samaras N. A comprehensive survey of machine learning methodologies with emphasis on water resources management. Appl Sci. 2023;13:12147.
19. Adnan MN, Islam MZ. A comprehensive method for attribute space extension for random forest. In: 17th International Conference on Computer and Information Technology (ICCIT). IEEE; 2014.
20. Amasyali MF, Ersoy OK. Classifier ensembles with the extended space forest. IEEE Trans Knowl Data Eng. 2013;26:549-62.
21. Uddin KMM, Dey SK, Babu HMH. A voice-assistive mobile application tool to detect cardiovascular disease using machine learning. Biomed Mater Devices. 2024;2:1246-57.
22. Marcelino P, de Lurdes Antunes M, Fortunato E, Gomes MC. Machine learning approach for pavement performance prediction. Int J Pavement Eng. 2021;22:341-54.
23. Zhang L, Jánošík D. Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches. Expert Syst Appl. 2024;241:122686.
24. Kalasampath K, Spoorthi K, Sajeev S, et al. Applications of explainable artificial intelligence (XAI): a review. IEEE Access. 2025.
25. Retzlaff CO, Angerschmid A, Saranti A, et al. Post-hoc vs ante-hoc explanations: XAI design guidelines for data scientists. Cogn Syst Res. 2024;86:101243.
26. Sadeghi Z, Alizadehsani R, Cifci MA, et al. A review of explainable artificial intelligence in healthcare. Comput Electr Eng. 2024;118:109370.
27. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765-74.
28. Apley DW, Zhu J. Visualizing the effects of predictor variables in black-box supervised learning models. J R Stat Soc B. 2020;82:1059-86.
29. Molnar C, Casalicchio G, Bischl B. Quantifying model complexity via functional decomposition for better post-hoc interpretability. In: ECML PKDD 2019 Workshops. Springer; 2020.
30. Molnar C, Schratz P. iml: Interpretable machine learning. R package version 0.10. 2020;1.
31. Grebely J, Dore GJ. What is killing people with hepatitis C virus infection? Semin Liver Dis. 2011;31:331-9.
32. Dore GJ. The changing therapeutic landscape for hepatitis C. Med J Aust. 2012;196:629-32.
33. Thomas DL. Curing hepatitis C with pills: a step toward global control. Lancet. 2010;376:1441-2.

DOI: 10.5455/medscience.2025.04.092

34. Wahib AA, Seif El Nasr M, Mangoud AM, et al. Liver function profile in PCR-RNA positive Egyptian HCV patients and controls. J Egypt Soc Parasitol. 2005;35:451-66.

35. Bazeed FB, Elsayed EH, Abd El-Aziz AA. Evaluation of aminotransferases and bilirubin in different treatment regimens for chronic HCV infection. J Pharm Biol Sci. 2016;11:20-6.

36. Martínez Herreros A, Sangro B, García Rodríguez A, Pérez Grijalba V. Albumin–bilirubin score as an indicator of improved liver function after DAA therapy. JGH Open. 2022;6:496-502.

37. Kostadinova L, Shive CL, Zebrowski E, et al. Soluble markers of immune activation during IFN-free HCV therapy. Pathog Immun. 2018;3:149-70.