machines

MDPI

Article

Explainable Machine Learning Approach for Hepatitis C Diagnosis Using SFS Feature Selection

Ali Mohd Ali [1], Mohammad R. Hassan [1], Faisal Aburub [2], Mohammad Alauthman [3], Amjad Aldweesh [4,*], Ahmad Al-Qerem [5], Issam Jebreen [6] and Ahmad Nabot [6]

[1] Communications and Computer Engineering Department, Faculty of Engineering, Al-Ahliyya Amman University, Amman 19328, Jordan
[2] Department of Business Intelligence and Data Analytics, University of Petra, Amman 961343, Jordan
[3] Department of Information Security, Faculty of Information Technology, University of Petra, Amman 961343, Jordan
[4] College of Computing and Information Technology, Shaqra University, Shaqra 11911, Saudi Arabia
[5] Computer Science Department, Faculty of Information Technology, Zarqa University, Zarqa 13110, Jordan
[6] Software Engineering Department, Faculty of Information Technology, Zarqa University, Zarqa 13110, Jordan
* Correspondence: a.aldweesh@su.edu.sa

Abstract: Hepatitis C is a significant public health concern, resulting in substantial morbidity and mortality worldwide. Early diagnosis and effective treatment are essential to prevent the disease's progression to chronic liver disease. Machine learning algorithms have been increasingly used to develop predictive models for various diseases, including hepatitis C. This study aims to evaluate the performance of several machine learning algorithms in diagnosing chronic liver disease, with a specific focus on hepatitis C, to improve the cost-effectiveness and efficiency of the diagnostic process. We collected a comprehensive dataset of 1801 patient records, each with 12 distinct features, from Jordan University Hospital. To assess the robustness and dependability of our proposed framework, we conducted two research scenarios, one with feature selection and one without. We also employed the Sequential Forward Selection (SFS) method to identify the most relevant features that can enhance the model's accuracy. Moreover, we investigated the effect of the synthetic minority oversampling technique (SMOTE) on the accuracy of the model's predictions. Our findings indicate that all machine learning models achieved an average accuracy of 83% when applied to the dataset. Furthermore, the use of SMOTE did not significantly affect the accuracy of the model's predictions. Despite the increasing use of machine learning models in medical diagnosis, there is a growing concern about their interpretability. As such, we addressed this issue by utilizing the Shapley Additive Explanations (SHAP) method to explain the predictions of our machine learning model, which was specifically developed for hepatitis C prediction in Jordan. This work provides a comprehensive evaluation of various machine learning algorithms in diagnosing chronic liver disease, with a particular emphasis on hepatitis C. The results provide valuable insights into the cost-effectiveness and efficiency of the diagnostic process and highlight the importance of interpretability in medical diagnosis.

Keywords: hepatitis C; data augmentation; feature selection; classification algorithms; machine learning; SHAP

Citation: Ali, A.M.; Hassan, M.R.; Aburub, F.; Alauthman, M.; Aldweesh, A.; Al-Qerem, A.; Jebreen, I.; Nabot, A. Explainable Machine Learning Approach for Hepatitis C Diagnosis Using SFS Feature Selection. *Machines* **2023**, *11*, 391. https://doi.org/10.3390/machines11030391

Academic Editor: Dan Zhang

Received: 9 February 2023
Revised: 9 March 2023
Accepted: 11 March 2023
Published: 16 March 2023

1. Introduction

Hepatitis C is a liver disease that affects millions worldwide [1–3]. Early diagnosis is vital for effective treatment, and using machine learning is becoming an essential tool [4–6]. The liver is crucial for normal body function, including aiding digestion, producing proteins and enzymes, removing toxins, and storing vitamins and minerals [7]. Hepatitis C is caused by a virus and can be potentially life-threatening, but many recover without intervention [8,9]. However, delayed treatment can result in cirrhosis and other complications.

The motivation behind this research stems from the pressing need for improved methods of diagnosing hepatitis C, a disease that poses a significant threat to public health worldwide. Early and effective treatment is crucial in preventing the progression of the disease to chronic liver disease and reducing the risk of severe health consequences such as liver cirrhosis and liver cancer. With the increasing prevalence of medical data, machine learning algorithms have become a popular tool for developing predictive models for various diseases, including hepatitis C. However, these traditional machine learning models often need more interpretability and transparency, making it challenging for practitioners to understand the reasoning behind the predictions and validate the models' reliability. This research addresses these concerns by evaluating the performance of explainable machine learning algorithms in classifying hepatitis C and examining the interpretability and transparency of their predictions. Additionally, the study seeks to improve the efficiency and reduce the cost of predictive diagnoses by combining feature selection and data augmentation techniques with machine learning algorithms, thus contributing to the advancement of the field of explainable machine learning.

The organization of this article is as follows: Section 2 provides a concise overview of the relevant literature pertaining to machine learning techniques for classifying hepatitis C disease. Section 3 elaborates on the methodology implemented in this study, encompassing a thorough description of the data utilized. Section 4 presents the study's results and subsequent analysis. Finally, Section 5 concludes with a summary of the findings and suggestions for future research.

2. Related Work

The liver is an essential organ that plays a critical role in removing toxins from the blood, aiding in digestion and metabolism. Any disruption in its function can lead to significant consequences in the body. The major types of liver diseases include hepatitis and cirrhosis, and early symptoms of liver disease may include nausea and fatigue. However, liver problems may not be diagnosed until they are advanced, and other symptoms may include itching, yellowing of the skin, and dark urine.

The diagnosis of liver disease typically involves a blood test that assesses the levels of alanine aminotransferase (ALT), aspartate aminotransferase (AST), or gamma-glutamic transferase (GGT). However, a liver biopsy may be needed to confirm the diagnosis and determine if treatment is necessary.

In [10], a researcher developed a system that integrates data mining with Decision Tree (DT) and fuzzy logic to explore and control hepatitis C virus infection. The study used Trapezoidal Fuzzy Number (TFN) to predict the disease's outcome and achieved a prediction accuracy of 98.1%, which is higher than DT's prediction accuracy of 92.5%. This study suggests that integrating data mining techniques with fuzzy logic and DT can be an effective approach to explore and control hepatitis C virus infection. The high prediction accuracy achieved using TFN indicates the potential of this approach in improving the accuracy of diagnosis and treatment of liver diseases.

The work in [4] analyzed the data of 4962 hepatitis C virus (HCV) patients in Egypt from 2006 to 2017 using machine learning techniques to identify the presence of esophageal varices, which is a common consequence of chronic liver disease. The study aimed to find a non-invasive approach to detect the existence of esophageal varices instead of using upper gastrointestinal endoscopy, which is a burdensome and unpleasant procedure for many patients. The study used 24 clinical laboratory variables and six well-known classifiers, namely Neural Networks (NNs), Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), and Bayesian Network (BN) to predict the presence of esophageal varices. The dataset was obtained from the Egyptian National Committee to Combat Viral Hepatitis, which is responsible for the national treatment program for viral hepatitis patients in Egypt, and the Ministry of Health oversees it. The study achieved an accuracy rate of 67.8%, 66.3%, 67.2%, 65.6%, 66.7%, and 68.9% using SVM, RF, C4.5, MLP, NB, and BN classifiers, respectively. These accuracy rates indicate that machine learning algorithms have the potential to detect the presence of esophageal varices using non-invasive approaches. The study suggests that machine learning techniques can be effective in predicting the presence of esophageal varices in patients with chronic liver disease. This non-invasive approach could potentially reduce the burden on endoscopy units and improve patient experience.

The work in [11] developed a machine learning algorithm for predicting Hepatocellular Carcinoma (HCC) in patients with HCV-related chronic liver disease. They used a collection of filtered input variables to obtain the best variable subset, and they employed three different classifiers: Logistic Regression (LR), Decision Tree (DT), and Classification and Regression Tree (CART). The study found that the accuracy levels achieved using LR, DT, and CART were 96%, 99%, and 95.5%, respectively. This suggests that machine learning algorithms have a high potential for predicting HCC in patients with HCV-related chronic liver disease, which could help in early diagnosis and personalized treatment plans. The study also highlights the importance of selecting the best subset of input variables to improve the accuracy of the prediction model.

In [8], the authors used machine learning techniques to predict outcomes of HCV based on viral nucleotides. They employed four different classifiers: Decision Tree (DT), Support Vector Machine (SVM), Naive Bayes (NB), and Neural Network (NN). They used processed features to predict the response to Interferon-alpha (IFN-alpha) and ribavirin (RBV) medication. The authors generated ten attribute weighting models using the basic dataset's 76 attributes. These models were generated using various properties such as Chi-square, Gini index, Deviation, Info-Gain, Info-Gain Ratio, SVM, PCA, uncertainty, relief, and rule. Next, they categorized the 11 characteristics using SVM, NB, NN, and DT and achieved an average accuracy of 85% in predicting the response to IFN-alpha and RBV medication. The study highlights the potential of using machine learning techniques for predicting the response to medication, which could help in personalized treatment plans for patients with HCV. The authors in [9] developed a machine learning-based prediction model for HCV using two different classifiers: Random Forest and K-Nearest Neighbors Algorithm. They used a dataset that included 668 cases of mild to moderate class 0 cirrhosis and 717 cases of class 1 cirrhosis. They used different features and characteristics to train the model. The goal was to identify the most efficient combination of characteristics to improve the accuracy of the model. Their study highlights the potential of using machine learning techniques for predicting HCV and improving patient outcomes.

While the Decision Tree algorithm used by [10] achieved the highest accuracy, the studies had limitations in terms of not using Shapley techniques to interpret the models. Shapley techniques can help explain how the features in the model contribute to the predictions and can provide insights into the underlying relationships between the features and the target variable. By using Shapley techniques, the researchers could have gained a better understanding of the importance of different features in predicting the presence of HCV in patients. Table 1 shows the summary of the related works.

Table 1. Summary of related works.

Author	Dataset	Classifier	Result Accuracy	Notes
Yehia Helmy et al. [10]	Laboratory examinations of HCV data in Egypt 2008–2012 200 samples	DT	92.50	No Metaheuristic No feature Selection No Shapley
		TFN	98.10	
Mohamed M. Ezz et al. [4]	Egyptian National Committee, under the supervision of the Ministry of Health 4962 samples	RF	66.30	No Metaheuristic Feature Selection: Filter Warper No Shapley
		C4.5	67.20	
		MLP	65.60	
		NB	66.70	
		BN	68.90	

Table 1. *Cont.*

Author	Dataset	Classifier	Result Accuracy	Notes
Mahmoud ElHefnaw et al. [11]	Egyptian National Committee for the Control of Viral Hepatitis Kasr Al-Aini Hospital 4423 samples	LR	96.00	No Metaheuristic Feature Selection: Variable Selection No Shapley
		DT	99.00	
		CART	95.50	
Chew XinYing [9]	HCV for Egyptian patients 1385 samples	kNN	47.35	No Metaheuristic No feature Selection No Shapley
		SVM	52.64	
		RF	50.72	
		NB	51.68	
		NN	46.87	
		Bagging	51.20	
		Boosting	50.24	
Heba Mamdouh Farghaly [1]	HCWs in Egypt 859 samples	NB	92.66	No Metaheuristic No feature Selection No Shapley
		RF	94.06	
		KNN	90.8	
		LR	93.01	

3. Materials and Methods

3.1. Dataset

The dataset used in this study was obtained from The Hospital of Jordan University and contained information about 1801 individuals tested for the hepatitis C virus. The data were collected using different types of blood tests that can be performed. The first type of test listed is the albumin blood (ALB) test, which measures the amount of albumin in the blood. Low albumin levels can indicate liver or kidney disease or another medical condition. The second test listed is the alkaline phosphatase (ALP) test, which measures the amount of ALP in the blood. ALP is an enzyme found in many body parts, and the test results are expressed numerically. The third test listed is the alanine transaminase (ALT) test, which assesses liver health by measuring the amount of ALT enzyme in the blood. The results are also expressed numerically. The fourth test listed is the AST (aspartate aminotransferase) test, which measures the amount of the enzyme in the blood. The BIL test is also listed, and its results are expressed numerically. The cholesterol levels test measures the amount of cholesterol and certain fats in the blood; its results are expressed numerically. The creatinine test measures creatinine levels in blood and urine; its results are expressed numerically. The gamma-glutamyl transferase (GGT) test measures the amount of GGT in the blood, and its results are expressed numerically. The total protein test (PROT) measures the total amount of two classes of proteins found in the fluid portion of the blood, and its results are expressed numerically. The table also includes age, gender, patient ID, and class (infected or uninfected), expressed numerically for age and patient ID and as binary values (male/female or infected/uninfected) for gender and class. The class label attribute is the dependent variable for the machine learning algorithms and divides the records into two categories: infected and uninfected. The attribute values were obtained from various medical tests. The dataset consists of 1801 instances and 13 attributes, including the class attribute. The attributes are split into two binary attributes and eleven numerical attributes. The HCV diagnosis dataset consists of 1801 patient records, out of which 294 patients are HCV positive, and the remaining patients are HCV negative. These data will be used to train and test the machine learning algorithms to predict the likelihood of a person being infected with the virus. Table 2 shows the features and meaning.

Our study calculated the correlation coefficients between different features using Pearson’s method. Pearson’s correlation coefficient is a commonly used method to measure the linear correlation between two variables, with values ranging from −1 to 1, where 1 indicates a perfect positive correlation, 0 indicates no correlation, and −1 indicates a perfect negative correlation. It measures the strength of the relationship and the direction of the relationship between two variables. In the HCV dataset, the correlation between features can provide insight into the relationship between various factors that influence HCV infection and its progression. For example, the correlation between age and HCV infection rate can help us understand if older individuals are more susceptible to HCV infection.

Table 2. Features' meaning in the dataset.

#	No. Features	Description	Types
1	Albumin Blood Test (ALB)	Measures the amount of albumin in your blood. Low albumin levels can indicate liver or kidney disease or another medical condition.	Numerical
2	Alkaline Phosphatase (ALP)	The test measures the amount of ALP in your blood. ALP is an enzyme found in many parts of your body. Each part of your body produces a different type of ALP.	Numerical
3	Alanine Transaminase (ALT)	It is an enzyme that mainly exists in your liver. An ALT blood test is often included in a liver panel and comprehensive metabolic panel; healthcare providers use it to help assess your liver health.	Numerical
4	AST (aspartate aminotransferase)	It is an enzyme found mostly in the liver but also in muscles and other organs in your body. When damaged cells contain AST, they release the AST into your blood.	Numerical
5	BIL test lab		Numerical
6	Cholesterol Levels	A cholesterol test is a blood test that measures the amount of cholesterol and certain fats in your blood. Cholesterol is a waxy, fat-like substance found in your blood and every cell of your body.	Numerical
7	Creatinine Test	This test measures creatinine levels in blood and/or urine. Creatinine is a waste product your muscles make as part of regular, everyday activity. Normally, your kidneys filter creatinine from your blood and send it out of the body in your urine.	Numerical
8	Gamma-glutamyl Transferase (GGT) Test	This test measures the amount of GGT in the blood. GGT is an enzyme found throughout the body, but it is mostly found in the liver. When the liver is damaged, GGT may leak into the bloodstream.	Numerical
9	PROT	The total protein test measures the total amount of two classes of proteins found in the fluid portion of your blood.	Numerical
10	Age	Numerical	Numerical
11	Gender	male, female	Binary (0,1)
12	Patient id	Numerical	Numerical
13	Class	infected or uninfected	Binary Range (0,1)

Additionally, the correlation between HCV viral load and liver function can help us understand how HCV progression affects liver function. Understanding the correlation between features in the HCV dataset can help us make better predictions and develop effective treatments for HCV. Figure 1 shows the correlation matrix of the HCV dataset.

The feature selection problem is an important aspect of machine learning. The goal is to reduce the dimensionality of the feature set to minimize error in predicting the class. This is especially important when the feature set consists of many variables, making it difficult to solve the problem numerically.

In feature selection, there are two steps: (a) identifying irrelevant features and (b) identifying redundant features. Irrelevant features are those that have no relationship between input and output features, while redundant features are those that have a high correlation with other attributes. Feature selection eliminates these irrelevant and redundant features as a pre-processing stage.

Feature selection methods measure the relevance and redundancy of the features. Relevance refers to the relationship between two attributes, and a feature selection algorithm retains the attributes with relevance between input and output features. Redundancy refers to the correlation between features, and any method must eliminate the features with a high correlation and select the attributes with a low correlation with other attributes.

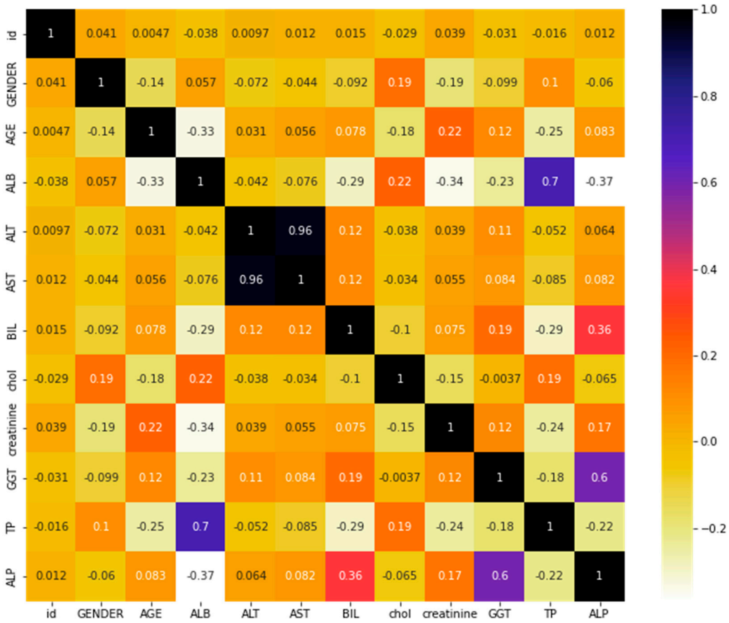


Figure 1. Simple correlation plot—HCV disease.

A feature f_i is considered relevant to the target concept Y if there is a subset S_i of features such that the probability of S_i excluding f_i is greater than zero, and the probability of Y given S_i and f_i is different from the probability of Y given S_i alone. If the feature f_i appears in every point representing Y, it is strongly relevant. If it appears in some points representing Y, it is weakly relevant. It is irrelevant if the feature does not appear in any point representing Y.

A feature f_i is relevant to the target concept Y if $P(Y|S_i, f_i) \neq P(Y|S_i)$ for some subset $\mathrm{S_i}$, where S_i is the set of all features except f_i, and $P(S_i = F - f_i) > 0$.

If feature f_i appears in every instance of Y, it is strongly relevant.

If feature f_i appears in some instances of Y, it is weakly relevant.

If feature f_i does not appear in any instance of Y, it is irrelevant.

In this study, we used Sequential Feature Selection (SFS). Sequential Feature Selection (SFS) is a well-known wrapper method used for feature selection in machine learning algorithms. SFS is an iterative algorithm that starts with an empty feature set and adds features one at a time to the set based on evaluating a classifier's performance on the training data. The classifier performance evaluation uses a performance metric, such as accuracy or F1 score, to determine the best feature set for a specific problem. SFS works by adding the feature that improves classifier performance most until a predetermined stopping criterion is met. This criterion can be the maximum number of features to be selected or a threshold for performance improvement. The selected features form the feature subset used in the final classifier.

The feature subset evaluation using SFS is computationally expensive as it requires training the classifier for each feature subset and evaluating its performance. However, SFS provides a flexible way of feature selection as it allows the user to specify different performance metrics and stop criteria from fitting the specific problem. It is more suitable for problems where the number of features is relatively small, and the individual features are expected to have a strong relationship with the target variable. As a result, SFS is a powerful tool for feature selection and can be combined with data augmentation techniques to further improve the performance of machine learning algorithms in classifying HCV or any other problem.

3.2. Proposed Framework

The methodology for explainable machine learning [12] using feature selection and data augmentation involves several key steps. The first step is data collection, where a real-world dataset of hepatitis C in Jordan was collected and used in this study. The data were collected from the Jordan University Hospital and consisted of 1801 samples with 13 attributes, including a class label attribute.

After data collection, the next step is data pre-processing, which involves cleaning the data, handling missing values, and converting categorical variables into numerical variables if necessary. The third step is feature selection, where techniques are used to determine the most relevant features for the prediction task. Sequential Forward Selection (SFS) Feature Subset Selection was used as a feature selection technique [13].

Data augmentation was then performed to overcome the limited data problem and improve the model's robustness. The SMOTE technique generated new synthetic samples from the existing data. The next step was model selection, where multiple algorithms were evaluated, including Decision Trees, Random Forests, Support Vector Machines, and Neural Networks.

Once a suitable machine learning algorithm was selected, the model was trained on the pre-processed and augmented data. This involved splitting the data into training and validation sets and using the training set to fit the model. The model was then evaluated on the validation set to assess its performance.

The final step was Explainable AI (XAI) [14], which is crucial in ensuring the transparency and interpretability of the model. XAI techniques were used to understand the model's decision-making process and identify the features that contribute the most to the predictions. This study used techniques such as feature importance and partial dependence plots to provide interpretable insights into the model. Figure 2 shows the schematic framework for the method used.

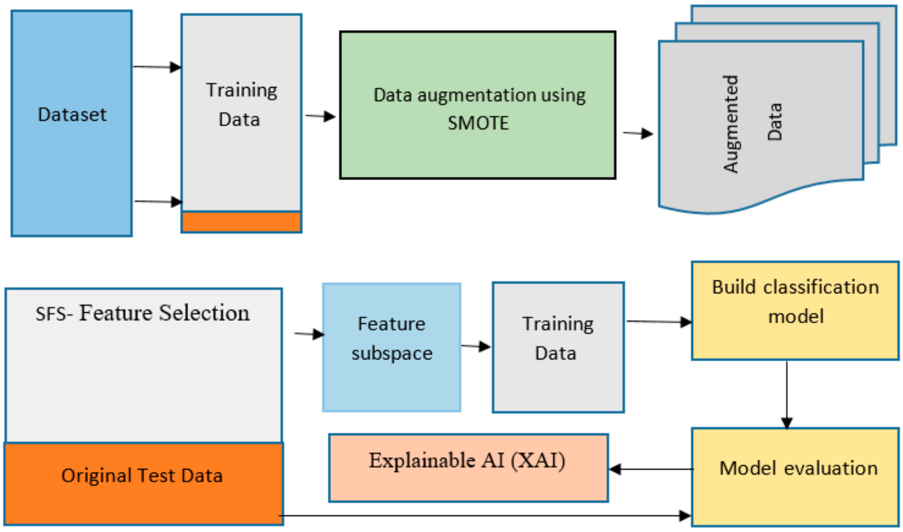


Figure 2. Framework for research methodology.

Explainable machine learning models are designed to explain their predictions through feature importance, decision rules, or visualizations, making the decision-making process transparent and understandable to human users [15,16]. This is crucial in applications where incorrect predictions, such as medical diagnosis or credit approval, could have severe consequences. Users can better understand the predictions and refine and improve the model over time by making it more interpretable, increasing trust in its outputs.

4. Experiments and Results

In this section, we will discuss the use of machine learning algorithms in classifying HCV by combining feature selection and data augmentation techniques. The main aim of this study is to improve the performance of machine learning algorithms by selecting the most important features and increasing the size of the dataset. In addition to feature selection, the study also employs data augmentation techniques, such as the synthetic minority oversampling technique (SMOTE), to increase the size of the dataset. SMOTE creates synthetic examples of the minority class by interpolating between existing examples. This technique is used to handle the imbalance in the dataset, where the majority class is overrepresented compared to the minority class.

- Linear Regression: A regression algorithm models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. The equation defines a straight line that can best approximate the relationship between the dependent and independent variables.
- Logistic Regression: A classification algorithm that models the probability of a discrete outcome, such as the occurrence of an event, given the values of one or more independent variables. The algorithm is an extension of linear regression and is useful for binary classification problems, where the outcome can only take two values, such as infected/uninfected.
- K-Nearest Neighbors: A non-parametric supervised learning algorithm for classification and regression. The algorithm classifies a new data point by finding its k-nearest neighbors in the training data and then taking a majority vote on the class labels of these neighbors.
- Random Forest: An ensemble learning method for classification and regression. The algorithm creates multiple decision trees by selecting random subsets of the training data and random subsets of the features. The final prediction is made by taking the average or majority vote of the predictions made by individual trees.
- Multi-Layer Perceptron: An artificial neural network used for classification. The network consists of input layers, hidden layers, and output layers. The hidden layers transform the inputs into outputs, and the network weights are adjusted during training.

The article discusses using various evaluation metrics such as precision, recall, AUC, and testing accuracy to evaluate ten machine learning classifiers, as shown in Table 3. The authors use the GridSearchCV method to tune the hyperparameters of the boosting-based classifiers to improve their accuracy.

Table 3. Evaluation measures.

Measure	Description	Equation
Accuracy	The percentage of correct predictions	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
Recall	The proportion of actual positives that are correctly identified	$Recall = \frac{TP}{TP+FN}$
Precision	The proportion of predicted positives that are correctly identified	$Precision = \frac{TP}{TP+FP}$
F-measure	The harmonic mean of precision and recall	$F = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$
R-squared	The proportion of the variance in the dependent variable that is predictable from the independent variables	$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i \left(y_i - \bar{y}\right)^2}$

In this study, 5-fold cross-validation is used to evaluate the performance of the machine learning models. This method partitions the data into five equally sized subsets, with four subsets used for training the model and the remaining subset used for testing. The main advantage of 5-fold cross-validation is that it provides a more reliable estimate of a model's performance than a single train/test split, as it uses all the available data for training and testing. It also helps to reduce the risk of overfitting, which can occur when a model is trained on a limited dataset. By repeating the process five times and using different subsets of the data for training and testing, cross-validation helps to ensure that the model generalizes well to new, unseen data. The results of the classification are based on the confusion matrix, which consists of the calculation of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values.

Table 4 summarizes results obtained by different classifiers under two combinations (data augmentation and feature selection). To make these results more readable, Figure 3a–d show the comparison between these classifiers regarding accuracy.

Table 4. Performance of different machine learning models.

		Without SMOTE				**SMOTE**			
	FS	Accuracy	Recall	Precision	F-measure	Accuracy	Recall	Precision	F-measure
LR	No	0.822	0.41	0.5	0.45	0.829	0.41	0.5	0.45
	Yes	0.829	0.41	0.5	0.45	0.823	0.41	0.5	0.45
KNN	No	0.831	0.92	0.51	0.46	0.831	0.92	0.51	0.46
	Yes	0.831	0.92	0.51	0.46	0.831	0.92	0.51	0.46
DT	No	0.820	0.41	0.49	0.45	0.820	0.41	0.49	0.45
	Yes	0.813	0.41	0.49	0.45	0.813	0.41	0.49	0.45
RF	No	0.822	0.57	0.51	0.43	0.824	0.51	0.58	0.76
	Yes	0.822	0.54	0.5	0.47	0.822	0.54	0.5	0.47
NN	No	0.829	0.41	0.5	0.45	0.822	0.41	0.5	0.45
	Yes	0.829	0.41	0.5	0.45	0.829	0.41	0.5	0.45

Table 4 compares the performance of 5 different machine learning models (Logistic Regression (LR), K-Nearest Neighbors (KNNs), Decision Tree (DT), Random Forest (RF), and Neural Network (NN)) for binary classification. The comparison is made between the results with and without the Synthetic Minority Oversampling Technique (SMOTE). The performance of each model is evaluated using four evaluation metrics: accuracy, recall, precision, and F-measure.

The results show that the performance of the models is not significantly impacted by the use of SMOTE, with some variations in accuracy, recall, precision, and F-measure. KNN shows the highest accuracy, recall, precision, and F-measure with 0.831, 0.92, 0.51, and 0.46, respectively, both with and without SMOTE. The lowest accuracy, recall, precision, and F-measure are shown by Decision Tree (DT), with 0.813, 0.41, 0.49, and 0.45, respectively, both with and without using SMOTE. The other models show intermediate results between KNN and DT.

The linear regression (LR) model's experimental results show that the SMOTE and feature selection accuracy decreased by 0.006 compared to using SMOTE without feature selection. The recall, precision, and F-measure criteria remained the same at 0.41, 0.5, and 0.45, regardless of whether feature selection was used. The RMSE decreased by 0.28 with feature selection and increased by 0.28 without feature selection. The AUC remained constant at 0.5 with or without feature selection. Without using SMOTE, the accuracy increased by 0.007 with feature selection and decreased by 0.007 without feature selection. The recall, precision, and F-measure criteria remained the same at 0.41, 0.5, and 0.45, regardless of whether feature selection was used.

The RMSE and AUC criteria remained constant at 0.41 and 0.45, respectively, with or without feature selection. The experiment's KNN method results showed that the performance metrics such as accuracy, recall, precision, F-measure, RMSE, and AUC remained the same with or without using SMOTE and feature selection. This indicates that the use of SMOTE and feature selection did not significantly affect the performance of the KNN model.

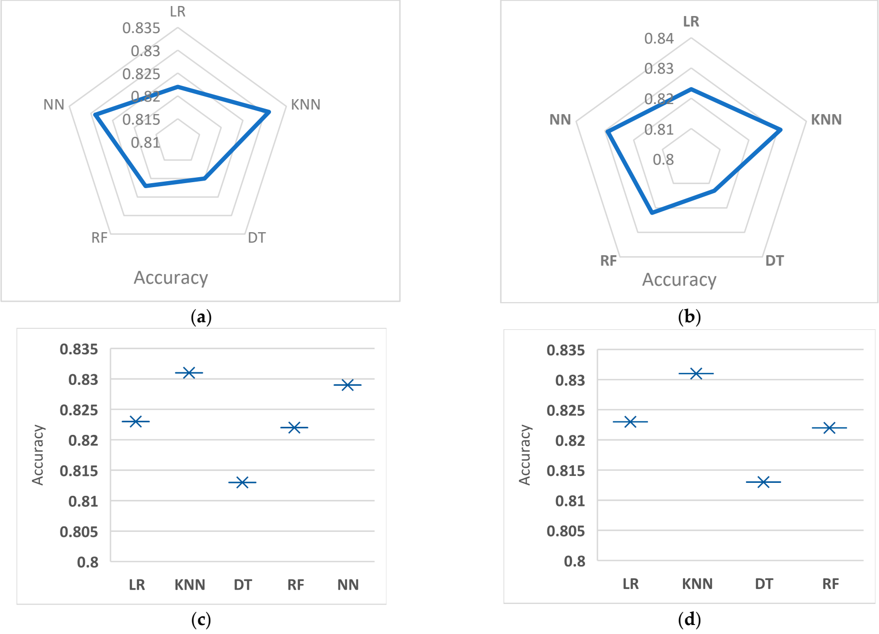


Figure 3. Accuracy under different combinations. (**a**) Feature selection (No), Augmentation (No). (**b**) Feature selection, (No) Augmentation (Yes). (**c**) Feature selection (No), Augmentation (Yes). (**d**) Feature selection (Yes), Augmentation (Yes).

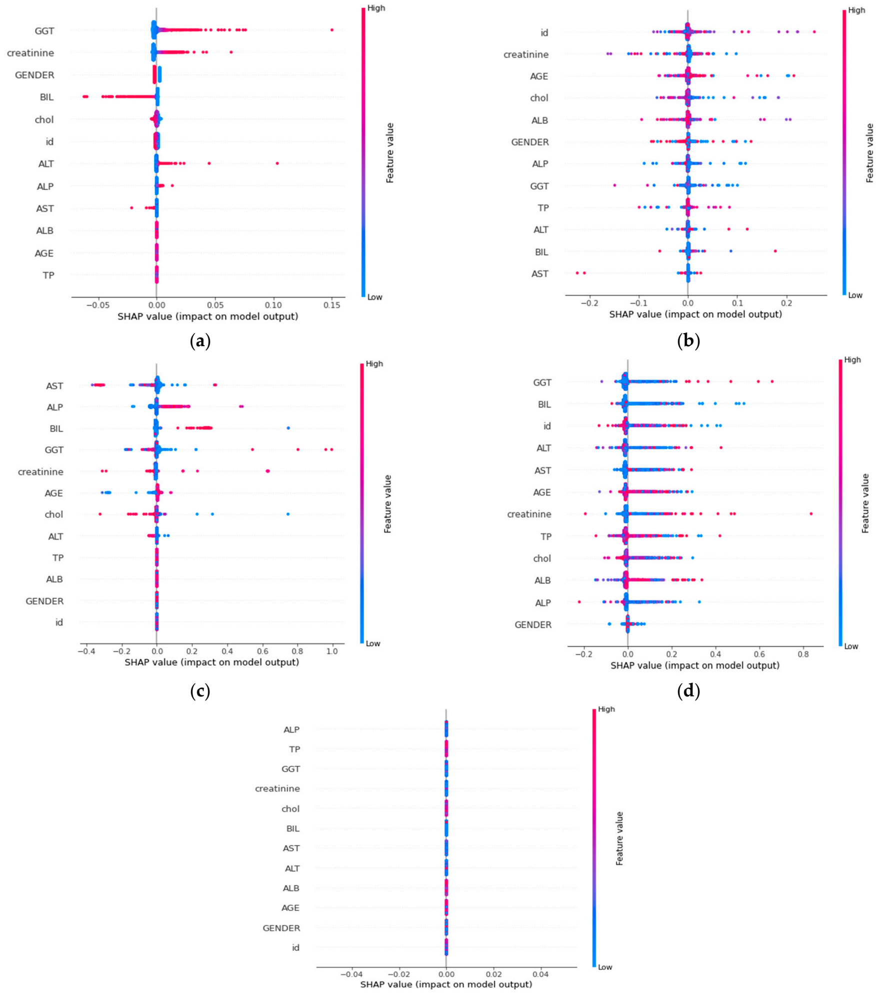
The experiment's Decision Tree method results showed that using SMOTE with feature selection did not significantly affect the accuracy, recall, precision, F-measure, RMSE, and AUC criteria compared to using SMOTE without feature selection. The differences in accuracy, RMSE, and AUC were only about 0.007 and 0.01, respectively, which can be considered a negligible change. The recall, precision, and F-measure values remained the same regardless of feature selection. The results of the experiment using Random Forest showed that the impact of using SMOTE with feature selection on accuracy was mixed, with a decrease of 0.002 but an increase of 0.002 without using feature selection. The recall criterion increased by 0.03 with feature selection but decreased by 0.03 without using it. The precision criterion decreased by 0.08 with feature selection and increased by 0.08 without using it. The F-measure criterion decreased by 0.31 with feature selection and increased by 0.31 without using it. The RMSE increased by 0.008 with and without feature selection. The AUC remained the same at 0.5 with and without feature selection. Without using SMOTE, the accuracy remained the same at 0.822, the recall decreased by 0.03 with feature selection and increased by 0.03 without using it, the precision decreased by 0.01 with feature selection and increased by 0.01 without using it, the F-measure increased by 0.04 with feature selection and decreased by 0.04 without using it, the RMSE remained the same at 0.421, and the AUC remained the same at 0.5. The experimental results with the Neural Network method indicate that using SMOTE with feature selection slightly increased the accuracy by 0.007 and decreased it without feature selection by 0.007. Other evaluation metrics such as recall, precision, F-measure, RMSE, and AUC remained unchanged with or without the use of feature selection and with or without the use of SMOTE. The results suggest that feature selection and SMOTE do not have a significant impact on the performance of the Neural Network method in this experiment.

SHAP (SHapley Additive exPlanations) values are a method for interpreting the output of machine learning models. They provide a way to understand how each feature contributes to the prediction for a given instance.

In the context of different classifiers, the SHAP values give an idea of the contribution of each feature toward the prediction made by that classifier. The SHAP values for each feature can be positive or negative, with positive values indicating that a feature contributes to the prediction being higher and negative values indicating that a feature reduces the prediction.

By aggregating the SHAP values for all instances, one can better understand the feature importance for the classifier. This can help interpret the predictions made by the classifier and identify the most important features that drive its behavior. SHAP values provide a way to get a deeper understanding of how classifiers work and can be useful in interpreting the results of different models.

The SHAP values were derived from the training data and showed how each feature contributed to the predicted outcome. Figure 4a–e shows SHAP values for different features at different classifiers. The study also used the SHAP values to visually interpret the model, making it easier for health practitioners or policymakers to understand and decide based on the results. Additionally, the study compared its results with state-of-the-art machine learning models for predicting hepatitis B and found that only one study considered model explainability, making this study an important contribution to the field.



(**a**) (**b**)

(**c**) (**d**)

(**e**)

Figure 4. SHAP values under different classifiers. (**a**): LR. (**b**): KNN. (**c**): DT. 5 (**d**): RF. (**e**): NN.

Table 5 compares the performance of different machine learning models in predicting hepatitis C disease from four research studies. Suiçmez et al. [17] achieved the highest accuracy of 98.7% using the Random Forest and multi-layer perceptrons. Dritsas and Trigka [18] obtained an accuracy of 80.1% using the voting classifier. Yağanoğlu [19] achieved an accuracy of 99.31% using various machine learning methods, pre-processing, and feature extraction. Saputra et al. and Safdari et al. [18] found that the Random Forest classifier had the best performance, with an accuracy of 97.29%. The proposed approach in the previous study applied Sequential Forward Selection (SFS) for feature selection and the SHapley Additive exPlanations (SHAP) method to explain the machine learning model's predictions, achieving an average accuracy of 83%. Overall, the results of these studies demonstrate the effectiveness of machine learning models in predicting hepatitis C disease, with some models achieving high accuracy rates.

Table 5. Comparison of different machine learning models in predicting hepatitis C disease.

Authors	Year	Dataset	Method	Accuracy (%)
Suiçmez et al. [17].	2023	Electronic health records of 615 patients, with 75 diseased individuals	Various ML algorithms with data mining techniques	98.7
Dritsas and Trigka [18]	2023	Liver disease dataset with 416 samples	Voting classifier with 10-fold cross-validation and SMOTE	80.1
Yağanoğlu [19]	2022	HCV dataset with added features and balanced with SMOTE	Various ML methods with pre-processing and feature extraction	99.31
Saputra et al. and Safdari et al. [20]	2022	HCV dataset, one with synthetic minority oversampling technique (SMOTE) and one without SMOTE	Six classification models (SVM, NB, DT, RF, LR, KNN) with Python programming language	97.29, 0.921, 0.963, 0.953, 0.972, 0.896, and 0.998 for accuracy, AUC, and different models, respectively
Proposed approach	2023	Dataset of 1,801 patients with 12 features	Various ML algorithms with SFS feature selection and SMOTE and SHAP for explainability	83

In analyzing HCV data collected from the Jordanian population, several important medical considerations exist for using machine learning in classification tasks. Firstly, ensuring that the data used for training and testing is representative of the population and has been collected systematically and ethically is important. This helps minimize potential bias in the model and ensures reliable and trustworthy results. Furthermore, it is important to consider the complexity of the HCV infection and the various factors that can affect its transmission, progression, and treatment. These factors can include demographic, lifestyle, environmental factors, and other comorbid conditions. A machine learning model should be able to account for these various factors and provide a robust and reliable estimate of the probability of HCV infection for a given patient.

Another important consideration is the choice of evaluation metrics for the model. In the context of HCV infection, it is important to balance the need for high accuracy with the need to minimize false positive or false negative predictions. This can help to ensure that patients who are infected with HCV receive appropriate treatment and those who are not infected are not subjected to unnecessary testing and treatment. Using machine learning for classifying HCV data collected from the Jordanian population requires careful consideration of several medical and ethical factors. These factors include the quality and representativeness of the data, the complexity of the HCV infection, and the choice of evaluation metrics. By considering these factors, it is possible to develop machine learning models that provide reliable and trustworthy predictions of HCV infection in the Jordanian population. Using machine learning for classifying HCV datasets collected from the Jordanian population has several benefits. One of the main advantages of using machine learning algorithms is handling large amounts of data and identifying patterns and relationships that would be difficult to detect using traditional statistical methods. This can lead to improved accuracy in the classification of HCV patients, which is critical in the medical field, where early diagnosis and treatment can have a major impact on patient outcomes.

Additionally, machine learning algorithms can handle missing data and deal with noisy or incomplete data, which is often a challenge in medical datasets. This means that the results from these algorithms can be more robust and reliable, leading to more informed decision making for healthcare providers. Finally, machine learning algorithms can be trained and optimized for different criteria, allowing for a customized approach tailored to the analysis's specific needs and goals. Using machine learning to classify HCV datasets collected from the Jordanian population can greatly improve the accuracy and efficiency of HCV diagnosis and treatment, ultimately leading to better patient outcomes.

5. Conclusions

The present study evaluated the performance of various machine learning (ML) algorithms for the classification of hepatitis C virus (HCV) using feature selection and data augmentation techniques. The timely and accurate diagnosis of HCV is critical for effective treatment and better patient outcomes; however, this task is challenging, time-consuming, and costly. Therefore, the use of ML algorithms has been explored to address these issues and reduce the cost of predictive diagnoses. In this study, data augmentation techniques such as the Synthetic Minority Oversampling Technique (SMOTE) were utilized to handle imbalanced data and increase the size of the dataset. The results of this study demonstrate that the combination of feature selection and data augmentation techniques can improve the accuracy of HCV diagnoses, reduce the cost of predictive diagnoses, and provide better diagnostic tools for the early detection of HCV. Moreover, the study addressed the growing concern about the interpretability of ML models in medical diagnosis. The Shapley Additive Explanations (SHAP) method was utilized to explain the predictions of the developed ML model, specifically for HCV prediction in Jordan. The findings presented in this study encourage further research in this area and contribute to the advancement of the field of medical diagnosis using ML algorithms.

Author Contributions: Conceptualization, A.M.A., M.A., I.J., A.N., M.R.H., A.A.-Q. (Ahmad Al-Qerem), A.A. (Amjad Aldweesh) and F.A.; methodology, A.M.A., A.A. (Amjad Aldweesh), M.A.; software, I.J., M.A, A.M.A. and A.A.-Q. (Ahmad Al-Qerem); validation, A.M.A., A.N., M.A. and I.J.; formal analysis, A.N, M.R.H., I.J., and M.A.; investigation, A.M.A., A.N., M.A., I.J. and M.R.H.; resources, A.M.A., A.N., M.A., I.J. and M.A; data curation, A.M.A., A.N., M.A., I.J. and M.R.H.; writing—original draft preparation, A.M.A., A.N., M.A., I.J., A.A. (Amjad Aldweesh) and A.A.-Q. (Ahmad Al-Qerem); writing—review and editing, A.M.A., A.N., M.A., I.J., A.A. (Amjad Aldweesh) and A.A.-Q. (Ahmad Al-Qerem); visualization, A.M.A., A.N., M.A., I.J., A.A. (Amjad Aldweesh) and A.A.-Q. (Ahmad Al-Qerem); supervision A.A.-Q. (Ahmad Al-Qerem); project administration, A.M.A.; funding acquisition, A.A. (Amjad Aldweesh) All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not Applicable.

Acknowledgments: Author Amjad Aldweesh would like to thank the Deanship of Scientific Research at Shaqra University for Supporting this Research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hashem, S.; Esmat, G.; Elakel, W.; Habashy, S.; Raouf, S.A.; Elhefnawi, M.; Eladawy, M.I.; ElHefnawi, M. Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis C patients. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *15*, 861–868. [CrossRef] [PubMed]
2. Kham-Kjing, N.; Ngo-Giang-Huong, N.; Tragoolpua, K.; Khamduang, W.; Hongjaisee, S. Highly Specific and Rapid Detection of hepatitis C virus using RT-LAMP-coupled CRISPR–Cas12 assay. *Diagnostics* **2022**, *12*, 1524. [CrossRef] [PubMed]
3. Abd Elminaam, D.S.; Elashmawi, W.H.; Ibraheem, S.A. HMFC: Hybrid MODLEM-Fuzzy Classifier for Liver Diseases Diagnose. *Int. Arab. J. E Technol.* **2019**, *5*, 100–109.
4. Abd El-Salam, S.M.; Ezz, M.M.; Hashem, S.; Elakel, W.; Salama, R.; ElMakhzangy, H.; ElHefnawi, M. Performance of machine learning approaches on prediction of esophageal varices for Egyptian chronic hepatitis C patients. *Inform. Med. Unlocked* **2019**, *17*, 100267. [CrossRef]
5. Shukla, N.; Angelopoulou, A.; Hodhod, R. Non-Invasive Diagnosis of Liver Fibrosis in Chronic Hepatitis C using Mathematical Modeling and Simulation. *Electronics* **2022**, *11*, 1260. [CrossRef]
6. Alauthman, M.; Aldweesh, A.; Al-qerem, A.; Aburub, F.; Al-Smadi, Y.; Abaker, A.M.; Alzubi, O.R.; Alzubi, B. Tabular Data Generation to Improve Classification of Liver Disease Diagnosis. *Appl. Sci.* **2023**, *13*, 2678. [CrossRef]
7. Hashem, S.; ElHefnawi, M.; Habashy, S.; El-Adawy, M.; Esmat, G.; Elakel, W.; Abdelazziz, A.O.; Nabeel, M.M.; Abdelmaksoud, A.H.; Elbaz, T.M. Machine learning prediction models for diagnosing hepatocellular carcinoma with HCV-related chronic liver disease. *Comput. Methods Programs Biomed.* **2020**, *196*, 105551. [CrossRef] [PubMed]
8. KayvanJoo, A.H.; Ebrahimi, M.; Haqshenas, G. Prediction of hepatitis C virus interferon/ribavirin therapy outcome based on viral nucleotide attributes using machine learning algorithms. *BMC Res. Notes* **2014**, *7*, 1–11. [CrossRef] [PubMed]
9. Nandipati, S.C.; XinYing, C.; Wah, K.K. Hepatitis C virus (HCV) prediction by machine learning techniques. *Appl. Model. Simul.* **2020**, *4*, 89–100.
10. Ali, M.M.R.; Helmy, Y.; Khedr, A.E.; Abdo, A. Intelligent Decision Framework to Explore and Control Infection of Hepatitis C Virus. In Proceedings of the International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018), Cairo, Egypt, 22–24 February 2018; Springer: Cham, Switzerland, 2018; pp. 264–274.
11. Mamdouh, H.; Shams, M.Y.; Abd El-Hafeez, T. Hepatitis C Virus Prediction Based on Machine Learning Framework: A Real-world Case Study in Egypt. *Knowl. Inf. Syst.* **2023**. [CrossRef]
12. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–7 December 2017; Volume 30.
13. Manikandan, G.; Abirami, S. A survey on feature selection and extraction techniques for high-dimensional microarray datasets. In *Knowledge Computing and its Applications: Knowledge Computing in Specific Domains: Volume II*; Springer: Singapore, 2018; pp. 311–333.
14. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.-Z. XAI—Explainable artificial intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [CrossRef] [PubMed]
15. Sghaireen, M.G.; Al-Smadi, Y.; Al-Qerem, A.; Srivastava, K.C.; Ganji, K.K.; Alam, M.K.; Nashwan, S.; Khader, Y. Machine Learning Approach for Metabolic Syndrome Diagnosis Using Explainable Data-Augmentation-Based Classification. *Diagnostics* **2022**, *12*, 3117. [CrossRef]
16. Obaido, G.; Ogbuokiri, B.; Swart, T.G.; Ayawei, N.; Kasongo, S.M.; Aruleba, K.; Mienye, I.D.; Aruleba, I.; Chukwu, W.; Osaye, F. An interpretable machine learning approach for hepatitis b diagnosis. *Appl. Sci.* **2022**, *12*, 11127. [CrossRef]
17. Suiçmez, Ç.; Yılmaz, C.; Kahraman, H.T.; Cengiz, E.; Suiçmez, A. Prediction of Hepatitis C Disease with Different Machine Learning and Data Mining Technique. In *Smart Applications with Advanced Machine Learning and Human-Centred Problem Design*; Springer: Singapore, 2023; pp. 375–398.
18. Dritsas, E.; Trigka, M. Supervised Machine Learning Models for Liver Disease Risk Prediction. *Computers* **2023**, *12*, 19. [CrossRef]
19. Yağanoğlu, M. Hepatitis C virus data analysis and prediction using machine learning. *Data Knowl. Eng.* **2022**, *142*, 102087. [CrossRef]
20. Saputra, T.A.N.; Arizona, K.I.; Andrian, M.R.; Kurniadi, F.I.; Juarto, B. Random Forest in Detecting Hepatitis C. In Proceedings of the 2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, Indonesia, 25–26 August 2022; pp. 299–302.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.