



Data-Driven Hepatitis C Treatment Optimization with Ensemble Machine Learning

Kazem Taghandiki^{1*}, Bashir Bashiri², Navid Maktabi³

¹ Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran.

² Tallinn University of Technology, Department of Chemistry and Biotechnology, Akadeemia tee 15, Tallinn, Estonia.

³ Razi Clinical Research Development Center, Guilan University of Medical Sciences, Rasht, Iran.

ARTICLE INFO

Article Type:

Original Research

Received: 2025.01.19

Revised: 2025.03.01

Accepted: 2025.04.14

Keyword:

Machine Learning
Hepatitis C
Intelligent Decision Support
Systems
Medical Data
Voting algorithms

*Corresponding Author:

Kazem Taghandiki

Email: ktaghandiki@tvu.ac.ir

ABSTRACT

Hepatitis C is a serious and potentially life-threatening liver disease that often remains asymptomatic for years, making early diagnosis challenging. If left undetected, it can lead to severe complications such as cirrhosis, liver failure, and hepatocellular carcinoma. Given the limitations of traditional diagnostic methods, which often struggle with accuracy and reliability, there is a growing need for advanced computational techniques to improve early detection and classification. This study introduces an ensemble machine learning approach to enhance the accuracy of Hepatitis C classification. The proposed method integrates multiple classifiers, including Naïve Bayes, Support Vector Machine, Decision Tree, and Linear Regression, using a majority voting mechanism. By addressing class imbalance and optimizing feature selection, this ensemble approach enhances predictive performance. Experimental evaluations using the Hepatitis UCI dataset demonstrate that the proposed method achieves a remarkable accuracy of 97.5%, outperforming individual models and conventional techniques. These findings underscore the potential of ensemble learning in medical decision support systems, providing a reliable tool for early diagnosis, risk assessment, and timely intervention.



1. Introduction

Hepatitis C is among the most critical diseases affecting the human liver, and if left unidentified, it may lead to human mortality [1]. The progression of Hepatitis C can be slow, and in some cases, the virus remains dormant in the body for 10 to 20 years without becoming active. Chronic Hepatitis C infection can lead to two major health issues: liver cirrhosis and liver cancer. Furthermore, the fact that a Hepatitis C vaccine is not yet available adds to the challenges [2]. Within six months, only 30% of patients naturally recover, while the majority develop a chronic viral infection. Infected individuals often do not exhibit symptoms until the disease becomes apparent and progresses. Fatigue, drowsiness, and dizziness may be mistaken for general tiredness or poor study habits, making the situation even more challenging [3]. Taking into consideration the difficulties in the early diagnosis of hepatitis, it is crucial to develop a tool that eases this challenge and improves patient outcomes.

For the early diagnosis of Hepatitis C, we should be able to classify the medical data. In recent years, due to the increasing volume of medical data, data classification has turned into a challenge for medical centers such as hospitals [4]. Considering the substantial amount of data and its inherent complexity, using computational methods and intelligent decision support systems can overcome this challenge [1].

Machine learning models are among the computational methods that can aid data classification, and therefore, facilitating the early diagnosis (prediction) would support medical staff in implementing measures to prevent the spread of disease. Thus, accurate prediction of Hepatitis C infection can reduce the risk of liver damage and other consequences by identifying and treating patients early. Additionally, machine learning models can help healthcare providers optimize the allocation of treatment resources, reduce treatment costs, and improve patient outcomes. Previous studies, such as those by Ma et al. [5], Ahammed et al. [6], Nandipati et al. [7], Mamdouh et al. [8], El-Salam et al. [9], Hashem et al. [10], Syafa'ah et al. [11], and Kashif et al. [12], have significantly contributed to the field by employing various classification algorithms, achieving notable accuracies.

Recent research has increasingly focused on utilizing ensemble machine learning methods, which have outperformed traditional algorithms in classification tasks [13]. Ensemble methods use multiple algorithms instead of a single algorithm to make final predictions based on the individual predictions of each model [13, 14]. Ensemble machine learning algorithms work based on the voting rules. The prediction that has the maximum votes is selected as the

final prediction; therefore, these algorithms are also called ensemble voting methods.

In this study, an ensemble voting method was applied to predict Hepatitis C status using the Hepatitis UCI dataset [15]. Data preprocessing techniques were first applied to enhance dataset features, followed by balancing algorithms to address class imbalances. Subsequently, four machine learning algorithms (Naive Bayes, Support Vector Machine, Decision Tree, and Linear Regression) were used to create the ensemble model. The final prediction was determined by the majority decision based on the votes of each model, demonstrating promising accuracy in Hepatitis C diagnosis.

Moreover, recent research, such as that of Zilouchian et al. [16], has focused on current and future diagnostics for Hepatitis C virus infection, emphasizing advancements in early detection methods and diagnostic techniques [16]. Additionally, Noviany et al. [17] proposed an interpretable machine learning approach using a voting-based LightGBM model to predict Hepatitis C virus NS5B inhibitor activity, highlighting the growing role of machine learning in improving Hepatitis C diagnostics and treatment prediction [17].

2. An overview of previous works

Many researchers have used ensemble algorithms on medical data for different purposes. Among these researchers, Rosly et. al.[18] evaluated the performance of both single and ensemble classification models. A comprehensive approach utilizing ensemble algorithms and Multi-Layer Perceptron (MLP) was employed to predict the status of Hepatitis C. The combination of these techniques resulted in a noteworthy model accuracy of 86.25%.

In contrast, Sachdeva et al. [19] employed single machine learning algorithms, specifically Support Vector Machine (SVM), Logistic Regression (LR), k-Nearest Neighbors (KNN), and Random Forest, for the detection of Hepatitis C. Addressing the imbalance issue in the Hepatitis dataset, they applied the Synthetic Minority Over Sampling Technique (SMOTE) combination method. Notably, the Logistic Regression algorithm emerged with the highest accuracy, reaching 93.18%.

The investigation by Alizargar et al. [2] extended its scope by considering six machine learning algorithms, namely SVM, KNN, LR, Decision Tree, XGBoost, and Artificial Neural Networks (ANN). This study involved the application of these algorithms to two datasets for predicting Hepatitis C. The analysis revealed SVM and XGBoost as effective tools for Hepatitis C prediction, with accuracy and AUC exceeding 80%.

In a study conducted by Huynh et al. [20], daily blood test data was utilized for Hepatitis C diagnosis, employing six machine learning algorithms: SVM, KNN, LR, Decision Tree, XGBoost, and ANN. The goal was to identify suitable models for Hepatitis C detection. Ultimately, the SVM and XGBoost algorithms exhibited the highest accuracy, achieving an impressive 96.8% among the tested algorithms.

Mamdouh et al. [8] contributed by developing four machine learning models, including Naive Bayes (NB), Random Forest (RF), KNN, and Logistic Regression, for predicting Hepatitis C using a dataset comprising 859 patients. The RF model emerged as the most effective algorithm, achieving an accuracy of 94.88%.

In the research conducted by Ma et al. [5], various classification models were presented for predicting Hepatitis C using data collected from 615 individuals. Notably, the XGBoost algorithm outperformed other models, demonstrating an accuracy of 91.56%.

Hashem et al. [10] took a comprehensive approach by applying multiple machine learning techniques to develop classification models for predicting advanced chronic Hepatitis C using clinical data. The developed models included decision trees, multi-variable linear regression, particle swarm optimization, and genetic algorithms. Performance evaluation utilizing Receiver Operating Characteristic (ROC) curves revealed accuracies ranging from 66.3% to 84.4% and area under ROC (AUROC) from 0.73 to 0.76. These algorithms successfully predicted advanced chronic Hepatitis C in patients with Hepatocellular Carcinoma (HCC), the most common type of liver cancer.

Zilouchian et al. [16] explored the current and future diagnostics for Hepatitis C virus infection, focusing on improving early detection and diagnostic accuracy. Their work emphasized the need for non-invasive and more sensitive diagnostic techniques to identify Hepatitis C infections, which is crucial in preventing disease progression. By reviewing various biochemical and biological diagnostic tools, they highlight advancements that enhance detection rates. However, while their research primarily addressed the diagnostic aspect, our study complements their work by utilizing machine learning models to predict Hepatitis C status using the Hepatitis UCI dataset. This approach introduces a different angle by applying computational techniques to the classification task, aiming to improve diagnostic accuracy through the integration of multiple machine learning algorithms.

Noviandy et al. [17] proposed an interpretable machine learning approach to predict Hepatitis C virus NS5B inhibitor activity using a voting-based LightGBM model and SHAP values. Their research highlights the growing role

of interpretable machine learning in drug discovery, especially in predicting the effectiveness of NS5B inhibitors for treating Hepatitis C. By focusing on treatment prediction, they aim to enhance decision-making in clinical settings. In contrast to their work, which is centered around drug prediction, the present study focuses on Hepatitis C diagnosis by applying an ensemble voting method that combines multiple machine learning models to classify the disease status. While both studies leverage machine learning techniques, the current research emphasizes improving diagnostic accuracy through ensemble-based predictions, offering a distinct approach to addressing the Hepatitis C challenge.

There has been research on predicting other diseases using classification algorithms as well. For instance, an article by Chao[21] diverged from Hepatitis C and focuses on predicting breast cancer utilizing machine learning algorithms such as SVM, LR, and Decision Tree.

3. Research Method

The approach presented in this article includes three main stages, as illustrated in Figure 1. To generate input data, the Hepatitis dataset generated by Carnegie Mellon University and provided by UCI [15] was utilized. The dataset is explained in section 3.1. In the second step, the Hepatitis dataset was enhanced qualitatively and quantitatively through preprocessing methods, and explained comprehensively in section 3.2. Finally, the voting algorithm was designed. The algorithm utilizes four sub-algorithms (Linear Regression, Support Vector Machine, Naive Bayes, and k-Nearest Neighbors) and addresses the selection of the best patient status, as elucidated in section 3.3.

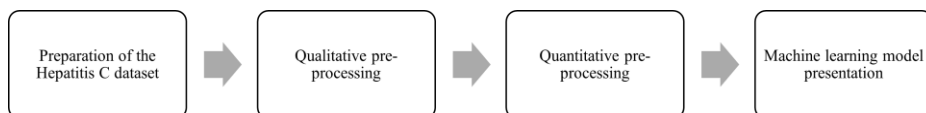


Figure 1. The general methodological framework followed in this research.

3.1. The Hepatitis dataset

The Hepatitis dataset, collected by Carnegie Mellon University for the development and testing of predictive models [15], holds relevance for classification and medical analysis purposes [18]. Comprising 19 features and test results from 155 individuals with Hepatitis, the dataset features medically significant attributes. These attributes can prove crucial for diagnosing liver diseases or providing insights into disease severity and treatment outcome predictions[22].

Primarily employed for testing classification algorithms such as decision trees, neural networks, and support vector machines [18], this dataset aims to

facilitate experiments predicting disease progression, determining treatment effectiveness, or evaluating medical risks [15]. Accuracy and dataset integrity are important; some features may have missing data that necessitate various methods to avoid deteriorating the accuracy of the final model.

The dataset for Hepatitis consists of 19 features and a target label indicating the status of patients (e.g., alive or deceased) [15]. The features, their explanation, and the type of data are provided in Table 1.

Table 1. Features included in the Hepatitis C dataset used in the current study

Feature	Explanation	Data Type
Age	Numerical value representing the age range of patients	Numerical
Sex	Categorization of patient gender	(male or female)
Steroid	Patient use of steroids for treatment	(yes or no)
Antivirals	Patient use of antiviral drugs	(yes or no)
Fatigue	Presence of fatigue in patients	(yes or no)
Malaise	Presence of discomfort, malaise, and fatigue in patients	(yes or no)
Anorexia	Presence or absence of anorexia in patients	(yes or no)
Liver Big	Is the liver larger than normal?	(yes or no)
Liver Firm	Is the liver tissue soft to touch or not?	(yes or no)
Spleen Palpable	Is the spleen larger than normal?	(yes or no)
Spiders	Presence of spider veins, a sign of liver disease	(yes or no)
Ascites	Presence of fluid in the abdominal cavity, a sign of liver disease	(yes or no)
Varices	Presence of varices in the stomach or esophagus	(yes or no)
Bilirubin	Bilirubin level in the blood, an indicator of liver damage	Numerical
Alkaline Phosphatase	An enzyme that can indicate liver damage	Numerical
Aspartate Aminotransferase	Another enzyme that can help assess liver damage	Numerical
Albumin	A protein produced in the liver, indicating the liver's functional status	Numerical
Prothrombin Time	Time taken for blood to clot, influenced by liver function	Numerical
Histology	Presence of pathological changes in liver tissue diagnosed through biopsy	Numerical

Finally, the feature "Target" serves as the dependent (label) feature in the dataset, indicating whether the patient is alive or not. All features were collected based on medical test results and observations. In some cases, information related to several features was not collected for certain patients, and this was improved in later sections using preprocessing methods.

3.2. Data preprocessing

The values of some features for hepatitis patients were not recorded in the dataset, leading to the presence of noisy data and a reduction in the overall accuracy of the classification model. Figure 2 illustrates the status of noisy features in the hepatitis dataset. According to Figure 2, approximately 15 features had noisy values, among which prothrombin, alkaline phosphatase, and albumin exhibit the highest levels of noise.

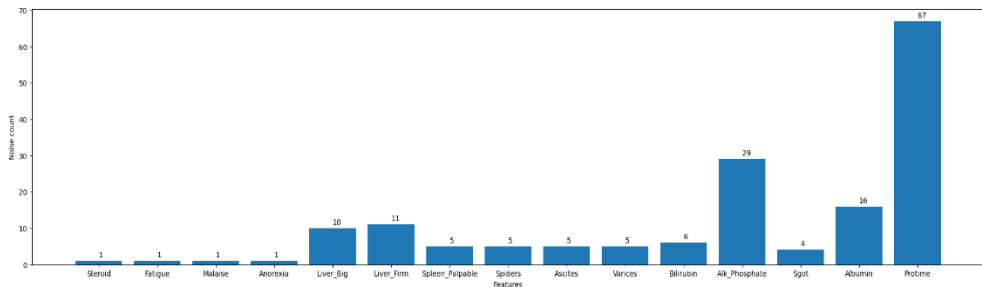


Figure 2. The status of noisy features in the dataset.

3.2.1 Improvement of Noisy Features

In the preprocessing stage of the hepatitis dataset, noise reduction was performed on the 15 features using the Pandas library, with specific methods applied based on the data type of the features. Reducing noise in the dataset is critical, as it helps to enhance the quality of the data, leading to improved performance of machine learning models. To address noisy features, two main strategies were employed: the Mean Method for numerical features and the Mode Method for nominal (categorical) features. These statistical techniques help reduce the impact of random variations, making the data more consistent and representative of the underlying patterns.

- 1. Mean Method for Numerical Features:** The mean method is a simple yet effective noise reduction technique for numerical features. By averaging the values of nearby observations, the means helps to reduce the influence of random fluctuations and outliers in the data. This is particularly useful when the dataset contains random noise or errors in numerical values, as the averaging process effectively neutralizes the effects of positive and negative noise, resulting in smoother and more stable data. This method functions by aggregating the values of contiguous observations and replacing noisy points with their mean value, leading to reduced variability and more reliable data representation[23, 24]. The Mean Method is beneficial in scenarios where the numerical features are affected by random noise and

variations that do not represent the actual data distribution. However, it is important to note that the mean method may not be suitable when the data has strong skewness, as the mean can be influenced by extreme values (outliers). Therefore, its effectiveness depends on the distribution of the data and the extent of the noise.

2. **Mode Method for Nominal Features:** For nominal or categorical features, the mode method is employed. The mode is the most frequently occurring value in the dataset and is particularly useful in handling discrete data with random noise or inconsistencies. In this approach, noisy or erroneous values in categorical features are replaced by the most frequent value (mode) in the dataset. This method is particularly effective in situations where measurement errors or inaccurate entries distort the data, as the mode represents the most likely or common occurrence in the dataset [23, 24]. One advantage of the Mode Method is that it does not require numerical computation and is ideal for categorical data, where other methods, such as averaging, are not applicable. However, it may not perform as well when the data has many equally frequent categories, as the mode may not be well-defined in such cases.

3.2.3 Resulting Impact on Dataset:

The use of the mean for numerical features and the mode for nominal features significantly reduced the noise in the dataset, improving its quality. After applying these methods, the data became smoother, and the models were less likely to be misled by random fluctuations or incorrect entries. It is noteworthy that noise reduction through these techniques aids in producing a more accurate representation of the underlying patterns in the dataset, which is crucial for the performance of subsequent machine learning models.

Furthermore, evaluating the performance of the model before and after applying noise reduction can provide valuable insights into how much the preprocessing steps contribute to improving the prediction accuracy. It would be interesting to compare the results of models trained on noisy data and clean data to assess the effectiveness of noise reduction.

3.2.4 Balancing the Dataset

One of the major challenges with the hepatitis dataset is its imbalanced class distribution. Among the 155 records in the dataset, only 32 belong to the "die" class, while the remaining instances are labeled as "live." This severe imbalance can lead to biased predictions from machine learning models, especially in the case of classifiers such as a voting classifier, which may tend to predict the

majority class ("live") more frequently, leading to reduced accuracy for predicting the minority class ("die"). Therefore, it is crucial to apply appropriate methods to balance the dataset, ensuring that the model learns from both classes more equally and can make fairer predictions.

To address this issue, several techniques for balancing class distributions are available, and scikit-learn offers a variety of options. Among the most commonly used techniques are as follows:

1. **Random Over-Sampling (ROS):** This technique involves generating additional samples for the minority class to increase its representation. While ROS helps balance the dataset, it may also lead to overfitting, as the model could become too familiar with the duplicated samples [25].
2. **Random Under-Sampling (RUS):** This method reduces the number of instances from the majority class, balancing the dataset by removing redundant samples. However, it may result in the loss of important information from the majority class, potentially harming the model's performance by reducing the overall size of the dataset [25].
3. **Combination of Over-Sampling and Under-Sampling:** This method combines both over-sampling of the minority class and under-sampling of the majority class, which often results in a more balanced and informative dataset. By applying both methods, this approach aims to minimize the disadvantages associated with each technique while preserving as much data as possible.

In the proposed approach, a combination of over-sampling and under-sampling was applied to balance the class distribution in the hepatitis dataset. Specifically, the number of "die" class instances increased to 83 through over-sampling, while the "live" class was reduced to 49 by under-sampling. This results in a balanced dataset with 132 records, significantly improving the class distribution and enhancing the fairness of the voting classifier's predictions.

To evaluate the effectiveness of this balancing process, a comparative analysis of model performance before and after balancing was conducted. As shown in Figure 3, the imbalance prior to balancing skewed the predictions towards the majority class, but while post-balancing, the model demonstrates improved fairness and accuracy, particularly in identifying the minority class ("die"). By addressing the class imbalance, the model is now better equipped to make more accurate predictions for both classes.

In future studies, it would be valuable to explore the impact of advanced techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or

ADASYN (Adaptive Synthetic Sampling Approach), which can generate synthetic samples in more sophisticated ways, potentially improving the generalization ability of the model without overfitting

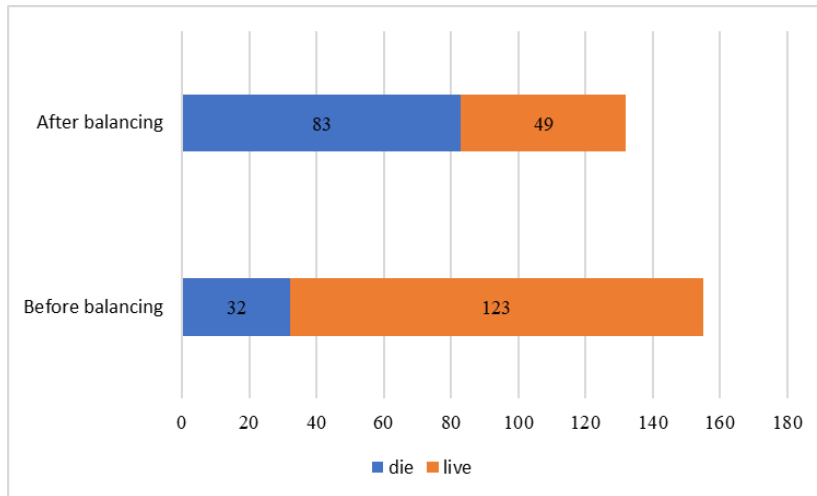


Figure 3. Pre- and Post-Balancing Results.

3.2.5 Transformation of Values

To build a Voting algorithm model, it is necessary to transform nominal (categorical) features of the hepatitis dataset into numerical features. For instance, the gender feature, which has values of male and female, should be transformed into a feature with values 0 and 1. Here, the value 1 represents being male, and the value 0 represents being female for a patient. Table 2 provides the data type of each feature in the hepatitis dataset before and after the process of transforming the values of nominal features. According to Table 1, bool data types, which include true and false values, have been transformed into int64 data types with values of 0 and 1. Feature number 20 corresponds to the target feature, and there is no need to transform it.

Table 2. Data Types of Features in Hepatitis Dataset Before and After Nominal Feature Value Transformation.

Before conversion			After conversion		
#	Column	Dtype	#	Column	Dtype
1	age	int64	1	age	int64
2	sex	object	2	sex	int64
3	steroid	bool	3	steroid	int64

4	antivirals	bool	4	antivirals	int64
5	fatigue	bool	5	fatigue	int64
6	malaise	bool	6	malaise	int64
7	anorexia	bool	7	anorexia	int64
8	liver_big	bool	8	liver_big	int64
9	liver_firm	bool	9	liver_firm	int64
10	spleen_palpable	bool	10	spleen_palpable	int64
11	spiders	bool	11	spiders	int64
12	ascites	bool	12	ascites	int64
13	varices	bool	13	varices	int64
14	bilirubin	float64	14	bilirubin	float64
15	alk_phosphate	float64	15	alk_phosphate	float64
16	sgot	float64	16	sgot	float64
17	albumin	float64	17	albumin	float64
18	protime	float64	18	protime	float64
19	histology	bool	19	histology	int64
20	target	object	20	target	object

3.3. Building the Voting Model

The Voting algorithm is an ensemble method in machine learning that combines predictions from multiple models to achieve better performance than any individual model [26]. In the context of classification problems, this algorithm adopts the "majority vote" approach, where each classifier casts a "vote" for a candidate. The class that receives the most votes is chosen as the final prediction [27].

Two types of voting exist:

a) Hard Voting:

In this scenario, the final prediction is based on the majority vote from the classifiers. Each classifier votes for a class, and the class that receives the most votes is selected as the final prediction [26].

b) Soft Voting:

In this case, the probability of each class as the output from each classifier is considered, and a weighted average of these probabilities is calculated to determine the final class. This method is typically more effective when classifiers produce reliable probability estimates [26].

3.3.1 Machine Learning Algorithms

Using four machine learning algorithms (linear regression, support vector machine, naive Bayes, and k-nearest neighbors), this research formed an ensemble algorithm for majority voting to provide the best decision. Each of these algorithms makes its prediction, and in the end, based on the majority vote, the final prediction is made by the voting algorithm. Figure 4 illustrates the execution process of the voting algorithm.

In this article, the training set and test set consisted of 80% and 20% of the hepatitis dataset, respectively. RL, NB, SVM, and KNN are machine learning algorithms used in designing voting algorithms. P1, P2, P3 and P4 are the predictions generated by the LR, NB, SVM and KNN, respectively. Finally, Pn represents the majority decision selected by the voting algorithm.

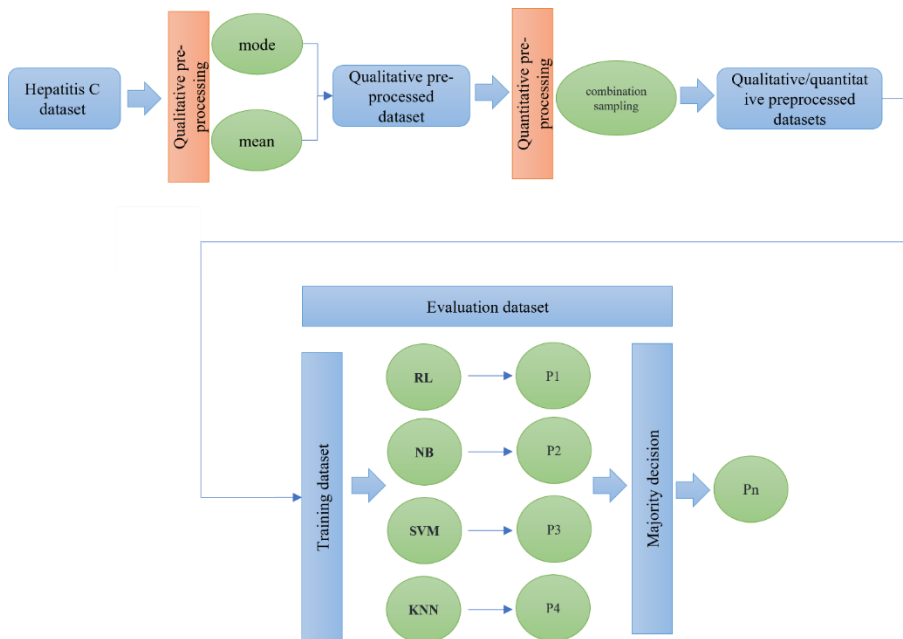


Figure 4. Visual Representation of Voting Algorithm Execution.

3.4. Evaluation of the performance of the ML algorithm

In the domain of binary classification, the confusion matrix stands as a fundamental tool, systematically evaluating the performance of a model by categorizing predictions into four distinct outcomes: True Positives (TP), True Negatives (TN), False Positives (FP), and False

Negatives (FN). TPs signify instances where the model correctly identifies positive cases, while TNs represent accurate identifications of negative cases. FNs indicate instances where positive cases are incorrectly labeled as negative, constituting a Type II error, and FPs occur when negative cases are incorrectly classified as positive, indicating a Type I error. These metrics are pivotal for a comprehensive understanding of the strengths and weaknesses of a classification model, particularly in terms of its accuracy in identifying positive and negative instances.

In the assessment of the proposed ensemble voting machine learning algorithm for predicting Hepatitis C status, a meticulous evaluation of key performance metrics sheds light on the effectiveness of the model in distinguishing between positive and negative instances. The confusion matrix, a cornerstone of binary classification, succinctly encapsulates the model predictions, laying the foundation for deriving essential metrics [28]. These metrics are listed in Table 3 in addition to their definition and formulas.

Table 3. Metrics, Definitions, and Formulas Derived from Confusion Matrix in Binary Classification.

Metrics	Formula	Description
Sensitivity	$\frac{TP}{TP + FN}$	Sensitivity, also known as the true positive rate, measures the proportion of actual positive cases that are correctly identified by a diagnostic test or classification model [29].
Specificity	$\frac{TN}{FP + TN}$	Specificity measures the proportion of actual negative cases that are correctly identified by a diagnostic test or classification model [28].
Precision	$\frac{TP}{TP + FP}$	Precision is the proportion of true positive results among the total positive results (true positives plus false positives) produced by a classification model [28].
Negative Predictive Value (NPV)	$\frac{TN}{TN + FN}$	Negative Predictive Value (NPV) is the proportion of true negative results among the total negative results (true negatives plus false negatives) produced by a classification model [28].
False Positive Rate	$\frac{FP}{FP + TN}$	The false positive rate is the proportion of actual negative cases that are incorrectly classified as positive by a diagnostic test or classification model [28].
False Discovery Rate	$\frac{FP}{FP + TP}$	The false discovery rate is the proportion of positive results that are false alarms among the total positive results produced by a classification model [28].

False Negative Rate		The false negative rate is the proportion of actual positive cases that are incorrectly classified as negative by a diagnostic test or classification model [28].
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Accuracy is the proportion of correct predictions (both true positives and true negatives) among all predictions made by a classification model [29].
F1 Score	$\frac{TP}{TP + \frac{1}{2}(FP + FN)}$	The F1 score is the harmonic mean of precision and recall, providing a single score that balances both measure [30].
Matthews Correlation Coefficient	$\frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$ The Matthews Correlation Coefficient (MCC) is a measure of the quality of binary classifications, considering true and false positives and negatives to produce a score between -1 and 1, where 1 indicates perfect prediction, 0 indicates a random prediction, and -1 indicates total disagreement between prediction and observation [23, 31, 32].	

4. Results and discussion

The confusion matrix for the proposed voting machine learning algorithm outlined in Table 3 identified patient outcomes as either "Live" or "Die." Among the total predictions, 10 instances were correctly identified as "Live" (True Positives), showcasing the proficiency of the algorithm in accurately recognizing patients who survived. Conversely, 1 case was falsely predicted as "Live" (False Positives), indicating instances where the algorithm incorrectly identified patients as survivors when they did not survive. Additionally, the algorithm exhibited only 1 False Negative, signifying its failure to recognize a surviving patient who was predicted to "Die." Lastly, the algorithm accurately predicted 29 cases as "Die" (True Negatives).

Table 3. The confusion matrix of the proposed voting machine learning algorithm.

		Actual Class	
		Live	Die
Predicated Class	Live	TP (10)	FP (1)
	Die	FN (1)	TN (29)

Comparison of key performance metrics before and after balancing the dataset is presented in Table 4. Before dataset balancing, the model demonstrated an accuracy of 85.1%, signifying correct classifications. The Sensitivity, emphasizing the proficiency of the model in identifying positive instances, was at 97.3%, attesting to its robust capability to minimize false negatives. Precision, assessing the accuracy of positive predictions, reached

86%, underscoring the precision of the model in positive classifications. The F1 Score, harmonizing Precision, achieved 91.3%, offering a comprehensive evaluation of the performance of the model.

Following dataset balancing, a notable improvement was observed in the overall correctness of models, with accuracy soaring to 97.5%. While there was a slight reduction, Sensitivity maintained a robust positive identification at 90.9%, critical for minimizing false negatives. Precision achieved perfection at 100%, denoting the absence of false positives among positive predictions. The balanced F1 Score, reaching 95.2%, underscored an elevated and well-rounded performance.

Balancing the dataset proved instrumental in refining the performance of the model, particularly in addressing class imbalances that could otherwise skew the learning process. The heightened Accuracy, Sensitivity, Precision, and balanced F1 Score underscore the efficacy of the proposed ensemble voting machine learning algorithm in predicting Hepatitis C status, positioning it as a valuable tool in medical decision support systems.

The results further delve into specific measures, both before and after balancing, to provide a nuanced understanding of the performance of the model. Prior to dataset balancing, the model showcased a Sensitivity of 97.3%, indicating its ability to accurately identify positive instances. However, after balancing, a slight reduction to 90.9% was observed, suggesting a nuanced change in positive identification.

Specificity, a measure of accurately identifying negative instances, exhibited a remarkable improvement from 33.3% before balancing to a perfect score of 100% after balancing. Precision achieved significant enhancement from 86% to a perfect 100%, emphasizing the model's precision in correctly classifying positive instances.

The Negative Predictive Value, indicative of the model's effectiveness in correctly identifying negative instances, improved from 75% before balancing to 96.6% after balancing. The False Positive Rate declined from 66.6% to zero after balancing, highlighting a substantial reduction in incorrectly classifying negative instances as positive.

The False Discovery Rate and False Negative Rate both reached zero after balancing, showcasing the model's improved accuracy in positive predictions and suggesting a nuanced change in the model's ability to recognize actual positive instances.

Overall, the dataset balancing played a crucial role in enhancing the performance of the model, resulting in a more accurate prediction of Hepatitis C status. The improvements in Sensitivity, Specificity, Precision, Negative

Predictive Value, False Positive Rate, False Discovery Rate, and False Negative Rate collectively contribute to the heightened accuracy and reliability of the model in clinical applications.

Table 4. Performance metrics of machine learning algorithms before and after balancing.

Measure	After balancing	Before balancing
Accuracy	0.975	0.851
Sensitivity	0.909	0.973
Specificity	1.000	0.333
Precision	1.000	0.860
Negative Predictive Value	0.966	0.750
False Positive Rate	0.000	0.666
False Discovery Rate	0.000	0.139
False Negative Rate	0.090	0.026
F1 Score	0.952	0.913
Matthews Correlation Coefficient	0.937	0.432

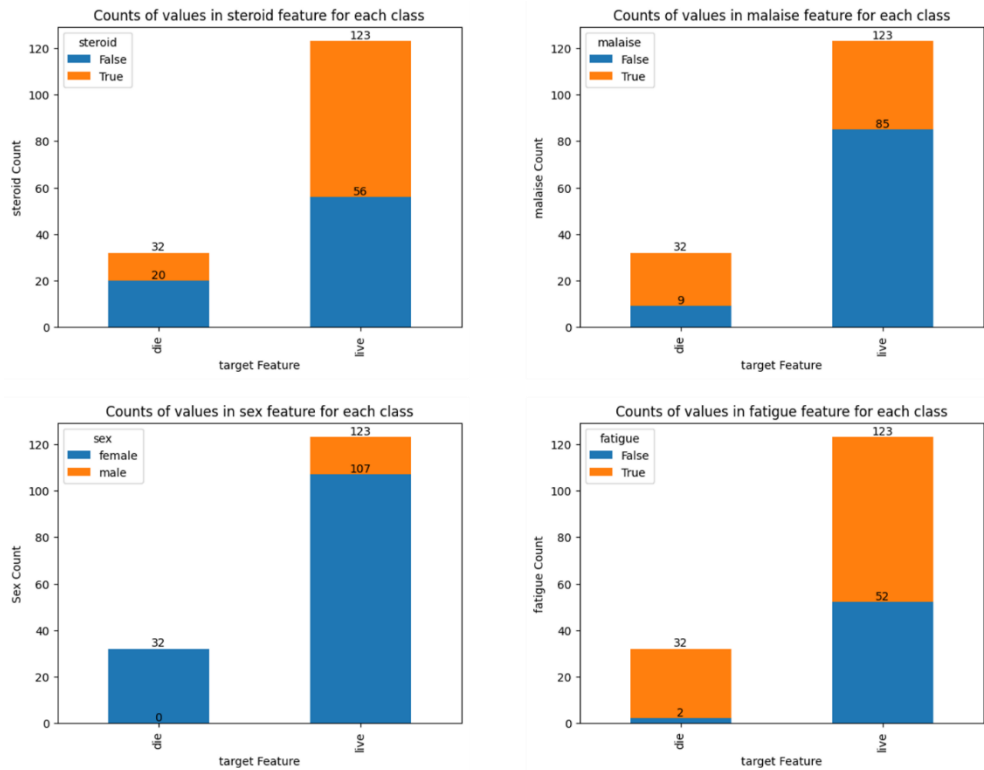


Figure 5. Distribution of values for the selected features: steroid, malaise, sex, and fatigue in the imbalanced database.

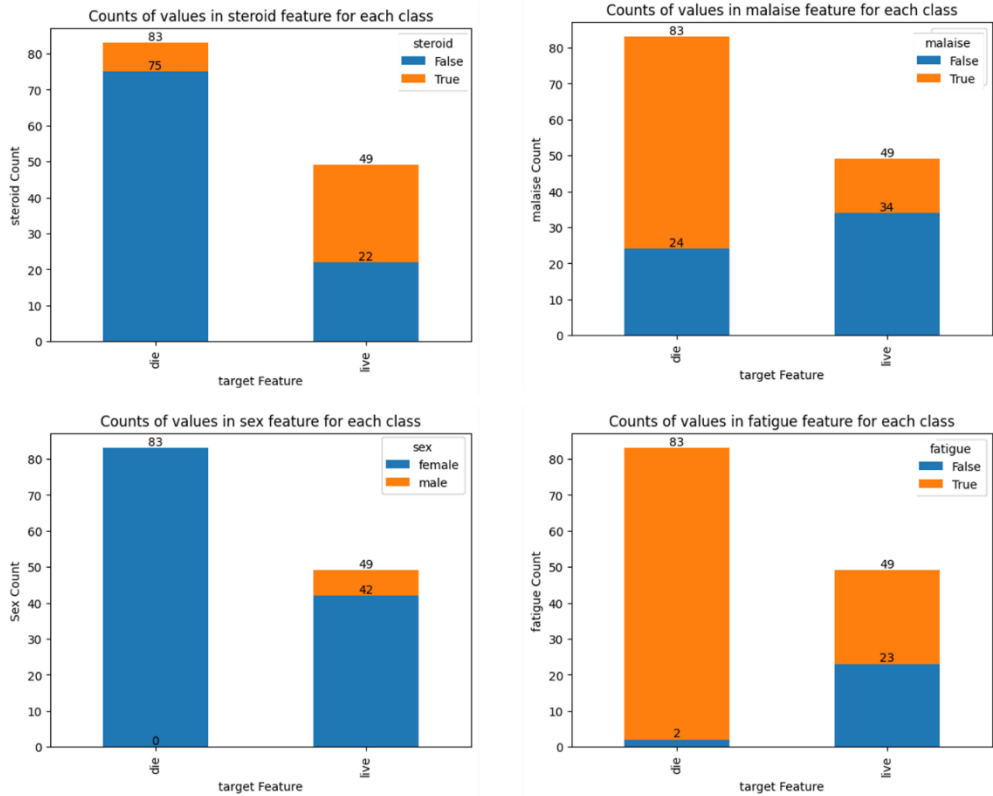


Figure 6. Distribution of values for the selected features: steroid, malaise, sex, and fatigue in the balanced database.

Figures 5 and 6 depict the distribution of values for the selected features: steroid, malaise, sex, and fatigue, concerning the class values in both imbalanced and balanced states of the hepatitis dataset. Therefore, the impact of employing combination sampling in balancing the dataset can be better understood. This algorithm enhances the representation of the minority target (Die) while diminishing it for the majority target (Live). As evident from Figures 5 and 6, the dataset balancing process plays a crucial role in providing a voting model with high accuracy.

The performance of the ensemble voting algorithm was investigated further by the Receiver Operating Characteristic (ROC) curve. The ROC curve illustrates the trade-off between the true positive rate and the false positive rate across different decision thresholds [2]. A higher area under the ROC curve value, such as the noteworthy 0.96 obtained in our model, signifies superior discriminative ability (Figure 7). It is important to note that an area under ROC value close to

1 is indicative of better model performance, underscoring the accuracy in correctly classifying positive instances while minimizing false positives [2].

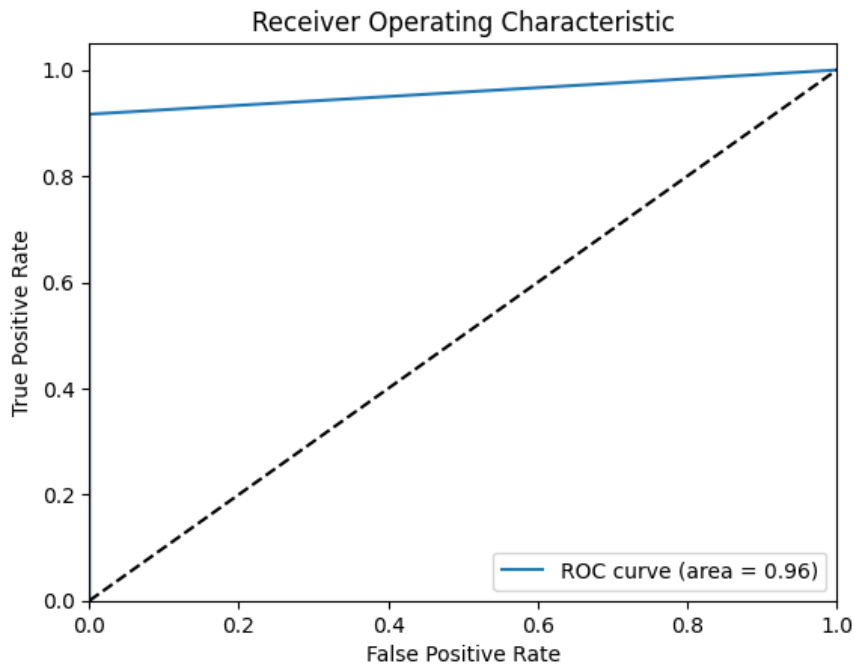


Figure 7. Receiver Operating Characteristic (ROC) curve.

In the present study's comparative analysis of the accuracy of various machine learning models, the proposed model is noticeable in achieving an accuracy of 97.5%. This performance is on par with the results reported by Rosly et al. [18], emphasizing the robustness and effectiveness of the proposed approach. Notably, this model surpasses the accuracies attained in other referenced studies, further underscoring its superior predictive capabilities. Figure 8 visually represents these findings, highlighting the commendable accuracy achieved by the proposed machine learning model in relation to other studies in the field.

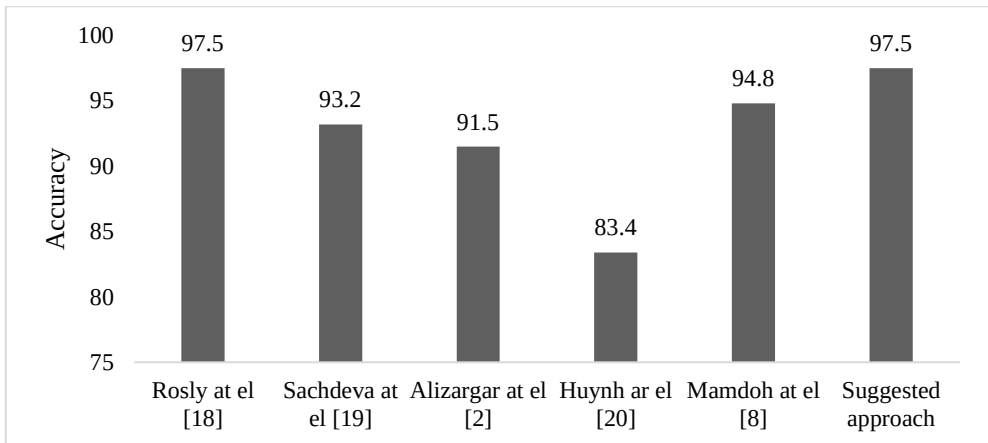


Figure 8. Comparative analysis of accuracy across the proposed algorithm and existing literature-based algorithms.

Previous research has explored various machine learning techniques for Hepatitis C diagnosis but has mostly relied on single classifiers such as Support Vector Machines (SVM) or Decision Trees, which may suffer from class imbalance or overfitting. The proposed approach leverages a majority voting ensemble method, which combines predictions from Naïve Bayes, SVM, Decision Tree, and Linear Regression, leading to more stable and accurate predictions.

Additionally, some studies focus solely on accuracy, whereas the present research incorporated multiple performance metrics such as specificity, precision, and F1-score to provide a more comprehensive evaluation of the model. These enhancements demonstrate the advantage of the proposed approach in real-world clinical applications, ensuring that misclassification rates are minimized while maintaining high sensitivity in detecting high-risk patients.

The significant improvement in the accuracy and robustness of the proposed model offers numerous real-world applications in medical decision support systems. By identifying high-risk patients with greater precision, the model enhances early diagnosis, allowing for timely and targeted interventions. It also optimizes resource allocation by minimizing misdiagnoses and reducing unnecessary medical procedures, leading to more efficient use of healthcare resources. Furthermore, by integrating into existing healthcare frameworks, the model supports clinical decision-making, helping doctors prioritize high-risk cases and improve patient outcomes.

5. Limitations of the Proposed Method

The ensemble model had several limitations. First, it demands higher computational power compared to single models, leading to increased processing overhead. Second, the performance of the model heavily relies on the dataset used, making it essential to validate its effectiveness across different Hepatitis datasets to ensure its generalizability. Lastly, while ensemble methods help mitigate overfitting, there is still a need to explore additional regularization techniques to further minimize this risk and enhance the robustness of the model.

6. Conclusion

Hepatitis C remains a major global health concern, with early and accurate diagnosis being essential for effective treatment and prevention of severe complications such as liver cirrhosis and cancer. Traditional diagnostic methods often suffer from limitations such as class imbalance, reduced accuracy, and the risk of misclassification. To address these challenges, this study proposed an ensemble machine learning model combining Naïve Bayes, Support Vector Machine, Decision Tree, and Linear Regression using a majority voting approach.

The proposed method significantly improves classification accuracy (97.5%), outperforming traditional single-classifier models. The ensemble approach effectively balances the dataset, reducing false positives to 0% and ensuring a high level of precision (100%) and specificity (100%). These improvements highlight the robustness and reliability of the model in real-world medical applications.

In conclusion, this study highlights the significant contributions of machine learning in the early detection, prognosis, and treatment planning of Hepatitis C. The proposed model not only enhances predictive accuracy but also supports better healthcare decision-making, ultimately leading to more effective and cost-efficient interventions. Future research should focus on refining these models further to improve their applicability and performance in clinical settings, ensuring their role in advancing healthcare practices.

7. References

1. Bajwa, J., et al., *Artificial intelligence in healthcare: transforming the practice of medicine*. Future healthcare journal, 2021. **8**(2): p. e188–e194. <https://doi.org/10.7861/fhj.2021-0095>
2. Alizargar, A., Y.-L. Chang, and T.-H. Tan, *Performance comparison of machine learning approaches on hepatitis C prediction employing data mining techniques*. Bioengineering, 2023. **10**(4): p. 481. <https://doi.org/10.3390/bioengineering10040481>

3. Hepatitis C. 2024; Available from: <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>.
4. Wang, L. and K. Zuo, *Medical data classification assisted by machine learning strategy*. Computational and Mathematical Methods in Medicine, 2022. **2022**(1): p. 9699612. <https://doi.org/10.1155/2022/9699612>
5. Ma, L., et al. *Prediction of disease progression of chronic hepatitis C based on XGBoost algorithm*. in *2020 International Conference on Robots & Intelligent System (ICRIS)*. 2020. IEEE. <https://doi.org/10.1109/ICRIS52159.2020.00151>
6. Ahammed, K., et al., *Predicting infectious state of hepatitis c virus affected patient's applying machine learning methods*. 2020 IEEE Region, 2020. **10**: p. 1371–1374. <https://doi.org/10.1109/TENSYMP50017.2020.9230464>
7. Nandipati, S.C., C. XinYing, and K.K. Wah, *Hepatitis C virus (HCV) prediction by machine learning techniques*. Applications of modelling and simulation, 2020. **4**: p. 89–100. https://arqiipubl.com/ojs/index.php/AMS_Journal/article/view/122
8. Mamdouh Farghaly, H., M.Y. Shams, and T. Abd El-Hafeez, *Hepatitis C Virus prediction based on machine learning framework: a real-world case study in Egypt: HM Farghaly et al*. Knowledge and Information Systems, 2023. **65**(6): p. 2595–2617. <https://doi.org/10.1007/s10115-023-01851-4>
9. Abd El-Salam, S.M., et al., *Performance of machine learning approaches on prediction of esophageal varices for Egyptian chronic hepatitis C patients*. Informatics in Medicine Unlocked, 2019. **17**: p. 100267. <https://doi.org/10.1016/j.imu.2019.100267>
10. Hashem, S., et al., *Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis C patients*. IEEE/ACM transactions on computational biology and bioinformatics, 2017. **15**(3): p. 861–868. <https://doi.org/10.1109/TCBB.2017.2690848>
11. Syafaah, L., Z. Zulfatman, and M. Lestandy, *Comparison of machine learning classification methods in hepatitis C virus*. JOIN (Jurnal Online Informatika), 2021. **6**(1): p. 73–78. <https://eprints.umm.ac.id/id/eprint/10873>
12. Kashif, A.A., et al., *Treatment response prediction in hepatitis C patients using machine learning techniques*. International Journal of Technology Innovation and Management (IJTIM), 2021. **1**(2): p. 79–89. <https://doi.org/10.54489/ijtim.v1i2.24>
13. Mohammed, A. and R. Kora, *An effective ensemble deep learning framework for text classification*. Journal of King Saud University-Computer and Information Sciences, 2022. **34**(10): p. 8825–8837. <https://doi.org/10.1016/j.jksuci.2021.11.001>
14. Feng, W., et al., *An ensemble machine learning approach for classification tasks using feature generation*. Connection Science, 2023. **35**(1): p. 2231168. <https://doi.org/10.1080/09540091.2023.2231168>
15. G. Gong, *Hepatitis*. 1983, UCI: <https://archive.ics.uci.edu/dataset/46/hepatitis>. <https://doi.org/10.24432/C5Q591>
16. Zilouchian, H., et al., *Current and future diagnostics for hepatitis C virus infection*. Chemosensors, 2025. **13**(2): p. 31. <https://doi.org/10.3390/chemosensors13020031>
17. Novianady, T.R., et al., *Interpretable machine learning approach to predict Hepatitis C virus NS5B inhibitor activity using voting-based LightGBM and SHAP*.

- Intelligent Systems with Applications, 2025. **25**: p. 200481. <https://doi.org/10.1016/j.iswa.2025.200481>
18. Rosly, R., et al., *Analyzing performance of classifiers for medical datasets*. International Journal of Engineering & Technology, 2018. **7**(2.15): p. 136–138. <https://doi.org/10.14419/ijet.v7i2.15.11370>
19. Sachdeva, R.K., et al., *A systematic method for diagnosis of hepatitis disease using machine learning*. Innovations in Systems and Software Engineering, 2023. **19**(1): p. 71–80. <https://doi.org/10.1007/s11334-022-00509-8>
20. Huynh, P.H. and V.H. Nguyen, *A novel ensemble of support vector machines for improving medical data classification*. Engineering Innovations, 2023. **4**: p. 47–66. <https://doi.org/10.4028/p-h0cef4>
21. Chao, C.-M., et al., *Construction the model on the breast cancer survival analysis use support vector machine, logistic regression and decision tree*. Journal of medical systems, 2014. **38**(10): p. 106. <https://doi.org/10.1007/s10916-014-0106-1>
22. Yildirim, P., *Filter based feature selection methods for prediction of risks in hepatitis disease*. International Journal of Machine Learning and Computing, 2015. **5**(4): p. 258–263. <https://doi.org/10.7763/ijmlc.2015.v5.517>
23. Zehtabvar, M., et al., *A review on the application of machine learning in gamma spectroscopy: Challenges and opportunities*. Spectroscopy Journal, 2024. **2**(3): p. 123–144. <https://doi.org/10.3390/spectrosci2030008>
24. Mishra, P., et al., *New data preprocessing trends based on ensemble of multiple preprocessing techniques*. TrAC Trends in Analytical Chemistry, 2020. **132**: p. 116045. <https://doi.org/10.1016/j.trac.2020.116045>
25. Batista, G.E., R.C. Prati, and M.C. Monard, *A study of the behavior of several methods for balancing machine learning training data*. ACM SIGKDD explorations newsletter, 2004. **6**(1): p. 20–29. <https://doi.org/10.1145/1007730.1007735>
26. Parhami, B., *Voting algorithms*. IEEE transactions on reliability, 1994. **43**(4): p. 617–629. <https://doi.org/10.1109/24.370218>
27. Weermeijer, J., et al., *Applying multiverse analysis to experience sampling data: Investigating whether preprocessing choices affect robustness of conclusions*. Behavior Research Methods, 2022. **54**(6): p. 2981–2992. <https://doi.org/10.3758/s13428-021-01777-1>
28. Akula, R., N. Nguyen, and I. Garibay. *Supervised machine learning based ensemble model for accurate prediction of type 2 diabetes*. in 2019 SoutheastCon. 2019. IEEE. <https://doi.org/10.1109/SoutheastCon42311.2019.9020358>
29. Kim, S.J., K.J. Cho, and S. Oh, *Development of machine learning models for diagnosis of glaucoma*. PloS one, 2017. **12**(5): p. e0177726. <https://doi.org/10.1371/journal.pone.0177726>
30. Peng, G., et al., *Reducing false-positive results in newborn screening using machine learning*. International journal of neonatal screening, 2020. **6**(1): p. 16. <https://doi.org/10.3390/ijns6010016>
31. Taghandiki, K., *Quantum machine learning unveiled: a comprehensive review*. Visual Journal of Technical and Vocational Education, 2024. **1**(2): p. 29–48. <https://doi.org/10.48301/jear.2024.446673.1021>
32. Chicco, D., M. Warrens, and G. Jurman, *The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification*

assessment. Ieee Access, 2021. **9**: p. 78368–78381.
<https://hdl.handle.net/10281/430460>