ORIGINAL ARTICLE

Integrated neural network and evolutionary algorithm approach for liver fibrosis staging: Can artificial intelligence reduce patient costs?

Ali Nazarizadeh,* Touraj Banirostam,* Taraneh Biglari,* Mohammadreza Kalantarhormozi,[†] Fatemeh Chichagi,[‡] Amir H Behnoush,[§] Mohammad A Habibi[¶] and Ramin Shahidi[‖]

*Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, [‡]Students' Scientific Research Center (SSRC), Tehran University of Medical Sciences, [§]Non–Communicable Diseases Research Center, Endocrinology and Metabolism Population Sciences Institute, Tehran University of Medical Sciences, Tehran, [†]The Persian Gulf Tropical Medicine Research Center, The Persian Gulf Biomedical Sciences Research Institute, Bushehr University of Medical Sciences, [‖]School of Medicine, Bushehr University of Medical Sciences, Bushehr and [¶]Clinical Research Development Center, Shahid Beheshti Hospital, Qom University of Medical Sciences, Qom, Iran

Abstract

Background and Aim: Staging liver fibrosis is important, and liver biopsy is the gold standard diagnostic tool. We aim to design and evaluate an artificial neural network (ANN) method by taking advantage of the Teaching Learning-Based Optimization (TLBO) algorithm for the prediction of liver fibrosis stage in blood donors and hepatitis C patients.

Methods: We propose a method based on a selection of machine learning classification methods including multilayer perceptron (MLP) neural network, Naive Bayesian (NB), decision tree, and deep learning. Initially, the synthetic minority oversampling technique (SMOTE) is performed to address the imbalance in the dataset. Afterward, the integration of MLP and TLBO is implemented.

Results: We propose a novel algorithm that reduces the number of required patient features to seven inputs. The accuracy of MLP using 12 features is 0.903, while that of the proposed MLP with TLBO is 0.891. Besides, the diagnostic accuracy of all methods, except the model designed with the Bayesian network, increases when the SMOTE balancer is applied.

Conclusion: The decision tree-based deep learning methods show the highest levels of accuracy with 12 features. Interestingly, with the use of TLBO and seven features, MLP reached an accuracy rate of 0.891, which is quite satisfactory when compared with those of similar studies. The proposed model provides high diagnostic accuracy, while reducing the required number of properties from the samples. The results of our study show that the recruited algorithm of our study is more straightforward, with a smaller number of required properties and similar accuracy.

Introduction

Chronic liver diseases are among the top causes of death worldwide, accounting for 2 million deaths worldwide annually, and are responsible for about 3.5% of total deaths. Half of this number can be attributed to complications due to cirrhosis and the remaining to viral hepatitis and hepatocellular carcinoma.[1] The hepatitis C virus (HCV), considered to be the main cause of liver cancer, is a blood-borne virus spreading via unsafe drug injection, blood transfusion, and unsafe sexual practices.[2] While 95% of patients are manageable, it can lead to liver cirrhosis or cancer. After 6 months, the chronic phase of the disease starts, influencing the liver and causing the production of inflammatory mediators, which makes the liver produce fibrous proteins to repair the damage.[2] This is called liver fibrosis, which results in less blood flow to the liver cells and eventually leads to necrosis.[3]

The METAVIR score has been designed to stage the fibrosis in chronic hepatitis C, with 0 representing no fibrosis and 4 showing severe cirrhosis or scarring.[4] However, this scoring system is based on liver biopsy, which is the gold standard for fibrosis staging. However, it has some drawbacks, which may limit its use. First, it is an invasive method of diagnosis, which can lead to several complications, including biliary peritonitis, hemorrhage, and even pneumothorax.[5] In addition, sampling errors and variations in histologic evaluations are inevitable.[6] Finally, biopsy cannot be used as a routine assessment tool for prognosis and follow-ups. Serum biomarkers and imaging

modalities are some of the alternatives for liver biopsy.[7] However, most of these options are associated with poor sensitivity and specificity: the aspartate aminotransferase-to-platelet ratio index (APRI) has 69% sensitivity and 77% specificity.[8]

Applications of artificial intelligence (AI) in medicine have grown rapidly in recent years.[9] AI can be used for prediction by classification methods using data mining and machine learning (ML) algorithms. Several AI projects have been conducted in order to predict the staging of hepatitis C, including some ML methods.[10–12] However, most of these have used a small number of parameters or could not achieve high accuracy.

Artificial neural network (ANN) is one of the most commonly used ML algorithms, which mimic the human brain in processing inputs and converting them to an output via some hidden layers.[13] The Teaching Learning-Based Optimization (TLBO) algorithm is a new population-based optimization algorithm that was proposed in 2011 and can find optimal solutions in a short period.[14,15] It is inspired by a number of learners taught by a teacher. Each learner is considered a solution, with its dimensions representing the parameters of the objective function of the given optimization problem.[16] Here we aim to formulate and investigate an ML algorithm on a pre-set dataset by utilizing ANN with TLBO optimization for feature selection to predict the staging of liver fibrosis in blood donors and hepatitis C patients.

Methods

Data mining and data source. Data mining is extracting data to analyze it more accurately and efficiently to improve decision-making efficiency. For our analysis, a dataset of blood donors and HCV patients obtained from the UCI machine learning repository[17] is used for our analysis.

Study outcome. The main outcome of the study was liver fibrosis staging, for which we compared different ML-based prediction models. In a random assignment process, the data were divided into train and test subsets (70% and 30%, respectively). Evaluation of the models was based on test data.

Data balancing. Because of the imbalanced data regarding positive and negative HCV subjects, we used the synthetic minority oversampling technique (SMOTE). It identifies the minority group's k-nearest neighbors, and then selects a set of neighbors that generates new data.[18] Moreover, the min-max method and substituting missing data with the mean were used to balance the dataset.

Teaching Learning-Based Optimization algorithm. We used TLBO[15] to reduce the number of features. This algorithm comprises two phases: (i) the Teacher phase, in which the learners get the knowledge from the teacher as the best solution available, and (ii) the Learner phase, in which learners learn through interacting with each other.[19,20] This method is based on a teacher's influence on students' output in a class, which means it works based on the mean. The teacher is the best way of selecting, and students' output is defined according to the mean of each selection method. In other words, this method compares the means of every random selection technique and chooses the

highest one. The most significant advantage of this method is finding the optimal solution in a short computational time.[14]

Machine learning algorithms. After data preparation, four algorithms were implemented in the balanced dataset: multilayer erceptron (MLP) neural network, Naive Bayesian, decision tree, and deep learning. Afterward, the models were compared. Our proposed method involved the integration of TLBO and MLP. All algorithms were implemented in the Keras environment.

Multilayer perceptron neural network. MLP is an extended neural network in which neurons are strongly connected. This study contains an input layer containing 12 neurons, an output layer with 5 neurons, and 1 hidden layer. We applied the function rule $[f(x) = \max\ (x,\ 0)]$ to enhance the network's learning pace. Furthermore, layer dropout was applied in the middle layer. In every election, only neurons with a high level of possibility (P-value >0.05) were used, thus preventing overfitting. In this study, the MLP technique used 70% of the dataset cases as training data and 30% as testing.

Naive Bayesian. This classification algorithm is based on the Bayes theorem with the objectivity assumption between the predictors. In this method, the dataset is considered as the input on which the analysis is conducted. Afterward, the class label will be predicted using Bayes's Theorem, and the probability of class in input data is calculated, which can help predict the class for the unknown data sample. This classification method can be especially suitable for large datasets.

Deep learning. Deep learning is a subset of ML techniques that focuses on training artificial neural networks with multiple layers, or "deep" networks. Unlike traditional ML models, deep learning models can automatically learn patterns and features from the data, making them suitable for complex tasks such as image recognition and natural language processing. In our research, deep learning was employed as one of the ML algorithms to enhance the accuracy of liver fibrosis staging predictions.

Decision tree. This is one of the most commonly used techniques for classifying algorithms. This algorithm's basis is to divide data into subcategories based on a series of questions. The starting point is the root node, which is the tree's root centered on the highest entropy and contains all samples. Each node is then split into secondary or leaf nodes in binary form or a multisplit. This model is like a tree structure that includes a group of nodes. It contains all decision nodes (splits node with the condition) and leaf nodes. The workflow of the dataset processing is illustrated in Figure 1.

Algorithm evaluation. The primary metric for evaluating the models was accuracy, which is defined as the number of correct predictions divided by the total number of samples. In addition to accuracy, we also report three essential metrics for a comprehensive evaluation of the model's performance:
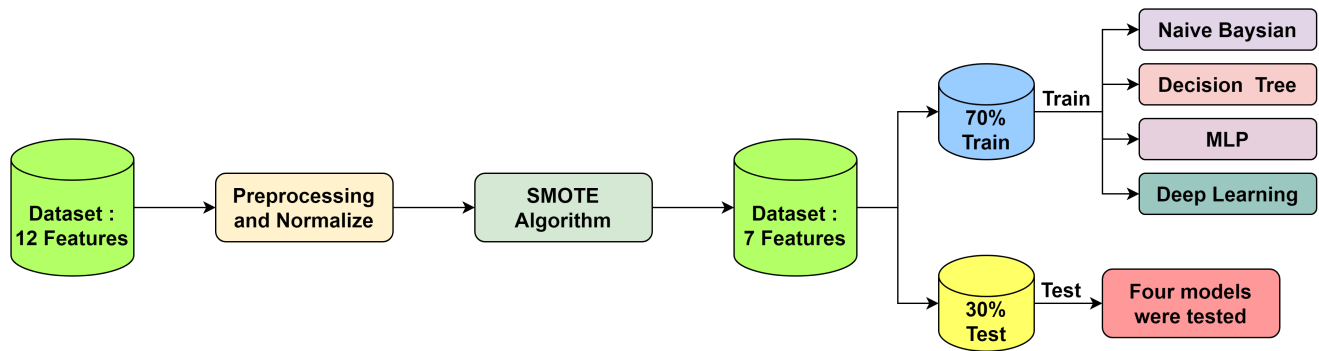
Figure 1 Workflow of dataset preprocessing and model training and testing processes.

Precision. This metric helps us to understand the proportion of true positive predictions (correctly identified liver fibrosis cases) out of all the positive predictions made by the model.

Recall. Also known as sensitivity or true positive rate, recall gauges the model's capacity to identify all actual positive cases. In the context of liver fibrosis, recall measures how effectively the model detects individuals with the condition, reducing the chances of false negatives.

F1 score. The F1 score is a balanced measure that considers both precision and recall, providing a single value that reflects the model's overall performance. It is particularly useful when there is an imbalance between positive and negative cases in the dataset. A high F1 score suggests a model that effectively identifies cases while minimizing false positives and false negatives.

Results

Our study focused on developing and evaluating ML-based models for predicting liver fibrosis stages in blood donors and hepatitis C patients. We employed a dataset from the UCI Machine Learning Repository, which included 615 records with an average age of 47.29 ± 9.993 years, all of which had the diagnosis of negative or positive HCV based on serology or histopathology: 75 patients with positive HCV test (including 24 hepatitis, 21 fibrosis, and 30 cirrhosis), each characterized by various clinical and demographic features. We removed seven cases because of suspicious diagnosis. Twelve data features were recruited for this study: age, sex, albumin blood test (ALB), alkaline phosphatase (ALP), alanine transaminase (ALT), aspartate transaminase (AST), acetylcholinesterase (CHE), cholesterol (CHOL), creatinine (Cr), bilirubin (BIL), gamma-glutamyl transferase (GGT), and proteins. The primary objective was to identify an optimized subset of features and assess the performance of ML algorithms in accurately predicting liver fibrosis stages. Table 1 provides a summary of patient characteristics in our dataset.

Feature selection and algorithm performance.
Our primary focus was to assess the performance of different ML algorithms in predicting liver fibrosis. We implemented the following algorithms: MLP neural network, Naive Bayesian,

Table 1 Summary of patient characteristics

Features	Range	Mean $\pm$ SD	Missing
Age	19–77	47.29 ± 9.993	0
Sex	—	237 Female/371 male	0
ALP	11.3–416.6	67.821 ± 25.2744	18
ALT	0.9–258	27.601 ± 21.2275	1
AST	12–324	34.369 ± 32.6224	0
ALB	20–82.2	41.819 ± 5.4067	1
CHOL	1.43–9.67	5.3788 ± 1.1194	10
Cr	8–1079	81.51 ± 49.721	0
BIL	1.8–25.4	11.474 ± 19.7706	0
GGT	4.5–650.9	38.244 ± 51.9532	0
CHE	1.42–16.41	8.2049 ± 2.1684	0
Proteins	51–90	72.253 ± 4.9323	1

decision tree, and deep learning. The models were compared based on their accuracy in predicting liver fibrosis stages.

In our study, we initially employed SMOTE to address the challenge of imbalanced data regarding positive and negative HCV subjects. SMOTE was applied to create synthetic instances in the dataset. The results of employing the SMOTE method are shown in Table 2. The introduction of SMOTE led to an improvement in diagnostic accuracy across all methods. However, the diagnostic accuracy of the Bayesian network model decreased after applying SMOTE, which was consistent with findings from previous studies using the same dataset. The highest accuracy was observed with deep learning and decision tree, while the Bayesian network model showed the lowest accuracy.

Subsequently, we used the TLBO algorithm to reduce the number of features from the initial set of 12. The selected features for our predictive models included age, ALB, ALP, ALT, AST, CHOL, and proteins.

The diagnostic accuracy of MLP algorithms trained with different features is provided in Table 3. The MLP model using all 12 features achieved an accuracy of 0.903, while the MLP model with TLBO-selected seven features maintained a high accuracy of 0.891. Notably, the reduction in features did not significantly compromise accuracy, indicating the effectiveness of feature selection in optimizing predictive performance. Additionally, we chose four blood indices at random to evaluate the

Table 2 Comparison of the diagnostic accuracy of models with and without using SMOTE

Algorithm	Accuracy without SMOTE	Accuracy with SMOTE
Naive Bayesian	0.870	0.855
Decision tree	0.897	0.962
MLP	0.902	0.956
Deep learning	0.908	0.966

Table 3 Diagnostic accuracy of MLP with different features

Algorithm	Number of features	Accuracy
MLP	12	0.903
MLP with TLBO (age, ALB, ALP, ALT, AST, CHOL, proteins)	7	0.891
MLP with random blood indices (GGT, proteins, Cr, CHOL)	4	0.587

power of random blood indices on liver fibrosis staging. Blood characteristics chosen at random included GGT, proteins, Cr, and CHOL. However, it is important to note that this method did not achieve a high level of accuracy in predicting the liver fibrosis stage.

Additionally, there are other useful algorithms such as support vector machine (SVM) and random forests, which are not reported in this paper because of their poorer accuracy or precision in our initial analysis of this dataset. Because of space limitations and the specific objective of our study, we just focused on the most accurate algorithms.

Discussion

Given the high importance of data mining in medical science, much research has been done in this field. In earlier research on the subject of liver complications, the integration of the TLBO's innovative algorithm has not been accomplished. Therefore, this study is quite innovative and not comparable to previous works. Commonly, in the proposed neural network model, the input data is 12 patient features, and the output consists of 5 stages of the liver fibrosis. However, aiming at our introduced algorithm, the number of features is reduced to seven inputs with preserved accuracy without significant change. The proposed algorithm reduces the volume of the problem and the cost required, and optimizes the proposed pattern. In addition to the primary approach of this study, which is the integration of the neural network and the TLBO's innovative algorithm, the dataset has been implemented with Bayesian algorithms and a decision tree. However, considering the accuracy of liver fibrosis staging in comparison to the previous studies discussed, it is reasonable to make a fair comparison between the results of our current study and those of the most recent prior research that achieved the highest diagnostic accuracy on this dataset.

This study deals with novel integration of neural networks and innovative algorithms. We used TLBO's innovative algorithm in our study. The integration is done in the pre-processing phase, and with the help of the TLBO algorithm, feature

selection has evolved. This study aimed to introduce a method by which we can predict the possibility of liver fibrosis with fewer blood tests and demographic data. Thus, we proposed a method using a combination of demographic data and blood tests, by which blood donors and patients infected with HCV could be staged for liver fibrosis without invasive procedures, and which could reduce the risks and costs. First, owing to the imbalance between healthy and infected cases, we implemented the SMOTE technique. Following this, we used different methods, among which decision tree and deep learning with 12 features produced the highest accuracy of 0.96. Moreover, the integration of TLBO and MLP methods, with a reduced set of seven primary features, yielded an accuracy rate of 0.89, while the MLP using all 12 features achieved a slightly higher accuracy of 0.903. Considering the preservation of the level of accuracy, this mentioned method with fewer features is specifically novel compared with previous ones.

In this study, we worked on the dataset that was used by some similar studies previously. An overview of the performance of models used in previous studies is provided in Table 4.

One study was performed using this dataset to recognize an efficient model of ML classification for diagnosing hepatitis C using 12 features. This study showed that the implementation of SMOTE improved the performance. In comparison with our model, those integrating SMOTE with NB, KNN, and logistic regression yielded accuracy rates of 0.84, 0.88, and 0.93, respectively.[23] In comparison, our study used the integration of the SMOTE with NB considering 12 features and reached an accuracy rate of 0.855.

Another study on this dataset was carried out using all 12 features by implementing the methods of random forest, MLP, and J48. In the mentioned study, 10-fold cross-validation method was implemented as a method for the evaluation of algorithms. Also, over-sampling and under-sampling methods were used to balance the dataset. Finally, the accuracy levels ranged from 0.65 to 0.99. With respect to MLP with 12 features, which is a common method between our study and the mentioned one, both studies produced about 0.95 of accuracy rate.[21]

In 2021, one study used the same dataset but using 10 features, by performing k-nearest neighbor (KNN), NB, neural network, and random forest. The highest accuracy rate was achieved for the neural network at 0.95, which is similar to our result.[22] Other methods reached accuracy rates between 0.89 and 0.94, which are not much different from our method. Considering TLBO implementation, our proposed method is not only more accurate but also can be considered more cost effective. Our proposed method is able to predict the liver fibrosis stage in blood donors and hepatitis C patients by measuring only seven clinical and demographic features instead of ultrasound-guided liver biopsy. Considering the costs and tests' risks, this method is safe, effective, economical, and time-saving for patients and hospitals compared with the conventional method of liver fibrosis stage detection.[24] As a result, fewer blood tests are more likely to be accepted by patients and insurance companies with regard to covering the costs.

Limitations and future direction. While our research does offer insights into hospital risks and costs, it is imperative to acknowledge our study's limitations. Considering the limited

Table 4 Comparison of the previous studies with our study

Title	Year	Method	Number of characteristics	Accuracy	Precision	Recall	F1 score
Machine-learning-based methods for handling imbalanced data in hepatitis diagnosis[21]	2021	MLP	12	Not reported	0.926	Not reported	0.929
Comparison of machine learning classification methods in hepatitis C virus[22]	2021	Decision trees	12	0.753	Not reported	Not reported	Not reported
Comparison of machine learning classification methods in hepatitis C virus[22]	2021	NN	12	0.951	0.88	0.82	Not reported
Machine learning models for diagnostic classification of hepatitis C tests[23]	2021	Naïve Bayes	12	0.849	Not reported	Not reported	0.841
Proposed method	2023	MLP	12	0.903	0.955	0.945	0.949
Proposed method	2023	Naïve Bayes	12	0.870	0.894	0.874	0.883
Proposed method	2023	Decision trees	12	0.962	0.965	0.955	0.956
Proposed method	2023	MLP with TLBO	7	0.891	0.869	0.89	0.879
Proposed method	2023	Deep learning	12	0.966	0.959	0.966	0.962

number of individuals in this study, an larger population size as well as various races and nationalities can affect the results in future studies. Moreover, 12 features are relatively high when compared with 615 records; we need algorithms with smaller numbers of features in future studies. Also, using various patient features in the prediction of other HCV outcomes, such as mortality, is highly recommended for future investigations. Also, we were unable to obtain information on liver biopsy for all patients, the specifics of the histology scoring methodology, and the platelet counts for calculating FIB-4 index despite our efforts, as well as lack of response from the dataset authors and relevant data on the source website. Finally, the lack of an acceptable external dataset prevented us from performing external validation in this study.

Conclusion

The methods of decision trees and deep learning showed the highest levels of accuracy with 12 features. Interestingly, with the use of TLBO and seven features, MLP reached an accuracy rate of 0.891, which is quite satisfactory and comparable to similar studies. The results between this study and previous studies were examined, which showed that the recruited algorithm of our study was more straightforward with fewer required properties and higher accuracy, which can lead to a diagnostic method with minimal cost and less required patient information.

Acknowledgments

References

1 Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. *J. Hepatol.* 2019; **70**: 151–71.

2 Lauer GM, Walker BD. Hepatitis C virus infection. *N. Engl. J. Med.* 2001; **345**: 41–52.

3 Besheer T, El-Bendary M, Elalfy H *et al.* Prediction of fibrosis progression rate in patients with chronic hepatitis C genotype 4: role of cirrhosis risk score and host factors. *J. Interferon Cytokine Res.* 2017; **37**: 97–102.

4 Pedrana A, Munari S, Stoové M, Doyle J, Hellard M. The phases of hepatitis C elimination: achieving WHO elimination targets. *Lancet Gastroenterol. Hepatol.* 2021; **6**: 6–8.

5 Piccinino F, Sagnelli E, Pasquale G, Giusti G. Complications following percutaneous liver biopsy. A multicentre retrospective study on 68,276 biopsies. *J. Hepatol.* 1986; **2**: 165–73.

6 Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. *Hepatology.* 1994; **20**: 15–20.

7 Myers RP, Benhamou Y, Imbert-Bismut F *et al.* Serum biochemical markers accurately predict liver fibrosis in HIV and hepatitis C virus co-infected patients. *AIDS.* 2003; **17**: 721–5.

8 Köksal İ, Yılmaz G, Parlak M *et al.* Diagnostic value of combined serum biomarkers for the evaluation of liver fibrosis in chronic hepatitis C infection: A multicenter, noninterventional, observational study. *Turk. J. Gastroenterol.* 2018; **29**: 464–72.

9 Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ.* 2019; **7**: e7702.

10 Tsvetkov V, Tokin I, Lioznov D. Machine Learning Model for Diagnosing the Stage of Liver Fibrosis in Patients with Chronic Viral Hepatitis C2021.

11 Akella A, Akella S. Applying Machine Learning to Evaluate for Fibrosis in Chronic Hepatitis C2020.

12 Butt MB, Alfayad M, Saqib S *et al.* Diagnosing the stage of hepatitis C using machine learning. *J. Healthcare Eng.* 2021; **2021**: 8062410.

13 Wesolowski M, Suchacz B. Artificial neural networks: theoretical background and pharmaceutical applications: a review. *J. AOAC Int.* 2012; **95**: 652–68.

14 Satapathy SC, Naik A, Parvathi K. Teaching Learning Based Optimization for Neural Networks Learning Enhancement. In: *Swarm, Evolutionary, and Memetic Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

15 Rao RV, Savsani VJ, Vakharia D. Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. *Comput.-Aided Des.* 2011; **43**: 303–15.

16 Baykasoğlu A, Hamzadayi A, Köse SY. Testing the performance of teaching–learning based optimization (TLBO) algorithm on combinatorial problems: Flow shop and job shop scheduling cases. *Inform. Sci.* 2014; **276**: 204–18.

17 Lichtinghagen R, Klawonn F, Hoffmann G. HCV data. UCI Machine Learning Repository. 2020. https://doi.org/10.24432/C5D612

18 Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002; **16**: 321–57.

19 Rao RV, Kalyankar VD. Parameter optimization of modern machining processes using teaching-learning-based optimization algorithm. *Eng. Appl. Artif. Intel.* 2013; **26**: 524–31.

20 Chen Y-K, Weng S-X, Liu T-P. Teaching–learning based optimization (TLBO) with variable neighborhood search to retail shelf-space allocation. *Mathematics*. 2020; **8**: 1296.

21 Orooji A, Kermani F. Machine learning based methods for handling imbalanced data in hepatitis diagnosis. *Front. Health Inform.* 2021; **2021**: 10.

22 Syafa'ah L, Zulfatman Z, Pakaya I, Lestandy M. Comparison of machine learning classification methods in hepatitis C virus. *J. Online Informatika.* 2021; **6**: 73.

23 Oladimeji OO, Oladimeji A, Olayanju O. Machine learning models for diagnostic classification of hepatitis C tests. *Front. Health Inform.* 2021; **2021**: 10.

24 Mendes LC, Stucchi RS, Vigani AG. Diagnosis and staging of fibrosis in patients with chronic hepatitis C: comparison and critical overview of current strategies. *Hepat. Med.* 2018; **10**: 13–22.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's website:

Data S1. Supporting information.