Informatics in Medicine Unlocked 17 (2019) 100267

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Performance of machine learning approaches on prediction of esophageal varices for Egyptian chronic hepatitis C patients

Shimaa M. Abd El-Salam [a,b,d], Mohamed M. Ezz [c,d], Somaya Hashem [a,b], Wafaa Elakel [e], Rabab Salama [e], Hesham ElMakhzangy [e], Mahmoud ElHefnawi [a,b,*]

[a] Department of Systems and Information, Engineering Research Division, National Research Centre, Egypt
[b] Bioinformatics Group, Centre of Excellence for Medical Research, National Research Centre, Giza, Egypt
[c] Department of Computer Science, College of Computer and Information Sciences, Jouf University, Saudi Arabia
[d] Department of Systems and Computers, Faculty of Engineering, Al-Azhar University, Cairo, Egypt
[e] Department of Endemic Medicine and Hepatology, Faculty of Medicine, Cairo University, Cairo, Egypt

ARTICLE INFO

Keywords:
Machine learning
Medical diagnosis
Esophageal varices
Hepatitis C virus
Prediction algorithms

ABSTRACT

Esophageal Varices is one of the most common side-effects of liver cirrhosis diseases which is detected by Upper endoscopy. Screening all patients implies many endoscopies will be needed, which increases the workload of endoscopy units. The aim of this study is to find solutions to diagnose the disease, by analyzing the patterns found in the data through classification analysis, using machine learning techniques for early prediction in cirrhotic patients based on their clinical examination. This research study attempts to propose a quicker and more efficient technique for disease diagnosis, leading to timely patient treatment. Our method analyzed 4962 patients with chronic hepatitis C from fifteen different centers in Egypt between 2006 and 2017. The dataset included twenty-four individual clinical laboratory variables. Esophageal Varices was present in 2218 patients and absent in 2,744 patients. Different types of feature selection (Filter-Wrapper) Approaches were applied to select the most significant features. The proposed model used six common algorithms including Neural Networks, Naïve Bayes, Decision Tree, Support Vector Machine, Random Forest and Bayesian Network to achieve our objective. The results showed that correlation and (p-value) based on filter method and Bayesian Network algorithm are well-suited for this analysis. Only nine variables: Gender, Platelet, Albumin, Total Bilirubin, Baseline_PCR, Liver, Spleen, Stiffness, and prothrombin concentration were the most significant predictors for Esophageal Varices. The Bayesian Network algorithm showed the highest performance; it achieved 74.8% and 68.9% for Area Under Receiver Operating Characteristic curves and accuracy, respectively. To conclude, machine learning techniques were able to predict Esophageal Varices in cirrhotic patients. The experimental results show that the Bayesian Network achieved better results than the other approaches.

1. Introduction

Every year, chronic liver disease leads to many complications that causes deaths globally. One of these complications is Esophageal varices, i.e., swollen blood vessels in the esophagus. Varices hemorrhage due to varicose rupture is the most common complication; it often occurs in patients with cirrhosis. Hence, predicting of presence the esophageal varices in cirrhotic patients is essential to avoid bleeding. Presently, the only diagnostic method for esophageal varices detection is via upper gastrointestinal endoscopy [1]. However, routine upper gastrointestinal endoscopy may be cost-ineffective since less than 50% of all patients with cirrhosis have varices. Therefore, checkup all cirrhotic patients with endoscopy mean many unnecessary endoscopies, which increase the workload of endoscopy units. Also, upper endoscopy can be an uncomfortable procedure for many patients. For this reason, the prediction of varices by noninvasive methods is essential to recognize patients who benefit from upper gastrointestinal endoscopy screening.

Machine learning (ML) is a branch of Artificial Intelligence (AI) and devoted to help make unguided choices by independently explicating input from the data. Machine learning algorithms acquire this data and

* Corresponding author. Department of Systems and Informatics, Engineering Research Division, National Research Centre, Egypt.
E-mail addresses: sa.awad@nrc.sci.eg (S.M. Abd El-Salam), maismail@ju.edu.sa (M.M. Ezz), wafaaelakel@link.net (W. Elakel), mahef@aucegypt.edu (M. ElHefnawi).

https://doi.org/10.1016/j.imu.2019.100267
Received 1 September 2019; Received in revised form 29 October 2019; Accepted 11 November 2019
Available online 12 November 2019

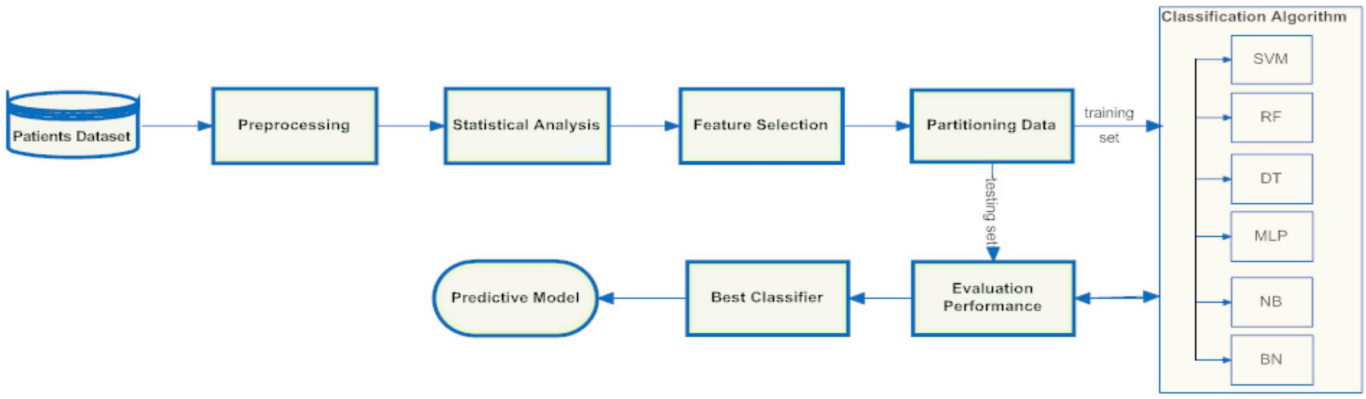

Fig. 1. Framework for predicting Esophageal Varices.

use it to build models for defining the actions taken by AI application. Data mining is a process that uses data analysis tools to discover knowledge and relationships within data to make valid predictions. Classification is an essential task in data mining and machine learning which aims to predict the classes of future data objects [2]. During the last few decades, researchers have applied machine learning algorithms in many fields to construct predictive models based on clinical data, such as predicting the response of therapy in Hepatitis C patients based on their clinical and biochemical information [3,4]. Also, the prediction of cirrhosis and composite fibrosis scores based on clinical data could be used as a non-invasive method for liver biopsy without adding financial expense to the treatment [5–8]. Some previous studies found that the Fibrosis-4 index (FIB-4) is a significant predictor. Stefanescu et al. [9] predicted the presence of varices using multiple regression, and concluded that esophageal varices could be predicted using FIB-4 with 62% Area Under the Curve (AUC) and 66.96% for accuracy. Sebastiani et al. [10] used multiple regression for predicting varices with 64% for AUC. Also, Bledar et al. [11] assessed FIB-4 as the only significant predictor of esophageal varices and variceal bleeding in Albanian patients; they achieved 66% for AUC. From the previous work, binary logistic regression was commonly used as a non-invasive method for predicting esophageal varices. Despite the large number of studies to date, no single or combination of non-invasive markers has been widely evaluated for predicting the esophageal varices in patients with cirrhosis [12].

Herein, six machine learning algorithms were proposed including Naïve Bayes (NB), Decision Tree (DT), Bayesian Nets (BNs), Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest (RF) for predicting the varices in chronic hepatitis C patients. The reason for selecting these six algorithms is because they are the most common algorithms that have been used with medical datasets to diagnose the diseases, and they outperform MLP in Refs. [3,4], DT in Refs. [3–5,7], Bayesian Nets in Ref. [13], and NB in Ref. [14]. Also, they had been commonly used for prediction of liver disease and demonstrated good success, as in Refs. [13,15,16]. Before building a prediction classifier, filter and wrapper feature selection approaches were applied to eliminate redundant features, so to reach a high performance [17]. The performance of the proposed algorithms was evaluated using the accuracy, precision, sensitivity, specificity analysis, and AUC. The rest of the paper is organized as follows. Section 2 introduces the dataset description, proposed Methods of classification techniques and evaluation. Section 3 presents the analysis and discussion of the results. Finally, conclusions are drawn in the last section.

2. Material and methods

This section is composed of three parts. Firstly, patient dataset description, preprocessing and feature selection to repair the data before applying data mining techniques is discussed. Secondly, the predictive models that applies a different DM algorithm to obtain the highest possible performance model is assessed. Thirdly, the evaluation parameters that evaluate the performance of each model and the best model which can predict the varices is determined. The framework of different parts for finding the best model is illustrated in Fig. 1.

2.1. Dataset description

A dataset of 4965 chronic C patients was analyzed for this study. The dataset has twenty-four individual clinical laboratory variables. This dataset was obtained from the Egyptian National Committee for Control of Viral Hepatitis in National Treatment Program of HCV patients in Egypt under the supervision of the Ministry of Health. Clinical and laboratory evaluation of the patients is a prerequisite for HCV treatment by Direct Antiviral agents (DAAs). The upper endoscopy was executed for all patients in a dataset in this study, with a comment of varices or no-varices. According to the endoscopic findings, patients were split into two main groups; the first consisted of 2747 (55.3%) patients with no-varices and the other had 2218 (44.7%) with varices. A collection of data was investigated and analyzed. The dataset variables include some blood serum descriptors such as Complete Blood Picture (CBC) that includes hemoglobin (HBGL), white blood cells (WBC) count, and platelet count, and liver function tests, including indirect bilirubin, alanine aminotransferase (ALT), albumin, aspartate aminotransferase (AST), total bilirubin, alpha-fetoprotein (AFP), international normalized ratio (INR), and prothrombin concentration (PC). In addition to age, gender, and body mass index (BMI), alcohol consumption and tobacco consumption were included in the patient questionnaire. Also, some laboratory tests were performed such as Diabetes, Baseline_PCR and Anti-Nuclear Antibody (ANA) creatinine, glucose and thyroid stimulating hormone test (TSH). Moreover, ultrasonography was applied to all patients, and the following details were recorded: Liver (cirrhotic, abnormal or normal), and Spleen (enlarged, average or removed). Transient Elastography was performed to measure liver stiffness (LS). Some variables were calculated in all patients such as AST-ALT ratio (AAR) and Fibrosis-4 index (FIB-4).

2.2. Preprocessing, statistical analysis and feature selection

Data quality is an important factor in the data mining process for disease prediction and diagnosis, because low-quality data may lead to inaccurate prediction results. In order to make our dataset more accurate for predicting varices, the preprocessing step was applied to clean the data. The cleaning process removes noisy data, handles null or missing values, and validates data type in the fields. Some values were deleted because they were noisy values or blanks; from the observation we found that some variables have a min value of 0, although the medical knowledge explains that such features cannot be zero.

Therefore, we considered these as missing values. Various methods have been suggested for handling missing values in datasets. In some cases, we replaced the missing values with the mean value, such as the BMI feature, and some others were calculated from equations such as Fib-4. Also, we applied normalization to some features, meaning transformation of all numerical values to fall within a small common range of [0,1] by calculating z-values, based on the mean and standard deviation. Each value in a variable is replaced by its Z-value; which is expressed by

$$\text{Z-value} = (\text{old value} - \text{mean}) / \text{standard deviation} \quad (1)$$

If it wasn't handled in this manner, these values can reduce the quality of our model result, and decrease the predictive power of the model, and the accuracy.

Statistical analysis methods were performed by using the MedCalc[1] [18] software tool. The continuous variables were presented as (mean value ± Standard Deviation). The Kolmogorov-Smirnov test was used for continuous variables with a normal distribution. The Mann-Whitney test was used for continuous variables with a non-normal distribution. The Chi-square test was utilized for categorical variables [19]. The score of the significance of the corresponding statistical test (p-value) was calculated [18]. The correlation coefficients were calculated between varices (target class) and all clinical variables. Correlation coefficients describe the strength of the relationship between variables and the varices group [19]. Pearson's correlation coefficient between two parametric variables is defined as the covariance of the two variables divided by the product of their standard deviations. The Spearman correlation coefficient is calculated for non-parametric variables. A correlation coefficient r gives values from −1 to +1, when the value is "±1" it means a total positive or negative correlation and when the value is "0" it mean no correlation between them.

Feature selection was used to select the significant attributes and eliminate redundant attributes. There are several types of feature selection methods [17]. In this paper, to select the significant attributes, we proposed six feature selection methods; three depend on the filter and the other three depend on the wrapper approach. In the filter approach, we used Correlation-based Feature Subset Selection (CFS) [20] with a score of significance (p-value); it is the probability that statistically significance is in fact zero (null hypothesis). These two methods have been used together, after calculating the score of significance for all attributes by (Chi-Square or Mann-Whitney) test, calculating the correlation coefficient between features and varices class. If less than the conventional 5% ($P < 0.0001$) it was considered as statistically significant; also if the value of correlation coefficient |r| was more than "0.10". Additionally, we applied other filter approaches including Information Gain (IG) and Principal Components Analysis (PCA). In the wrapper approach [17], we used greedy stepwise, Genetic algorithm (GA), and Particle Swarm Optimization (PSO). They were used to search through all possible subsets of attributes, and explore the mutual information between features.

2.3. The predictive models

For our study, ANN, Naïve Bayes, Bayesian Net, DT, (SVM), and Random Forest (RF) were applied to predict esophageal varices, because they were the most common classification algorithms that had been used for medical diagnosis and achieved better results. Our models took only a set of significant features depending on applying a filter approach; six different algorithms were applied using feature subsets to reach the best possible performance model. The dataset was split into 80% and 20% for the training and testing set respectively. After the learning process, the classifier will ready to recognize and classify the patients. Our model was able to identify the presence of the varices in patients depending on a set of clinical test analysis results; the next section explains the details.

SVM is used in the classification of linear and non-linear data. It transforms the original training data to a higher dimension using non-linear mapping. Within this new dimension, it searches for the linear optimal separating hyperplane. Data from two classes can be separated by hyperplane, with an appropriate nonlinear mapping to a sufficiently high dimension. Using support vectors and margins, the SVM finds this hyperplane. SVM implements the classification task by maximizing the margin that classifies both classes while minimizing the classification errors [14].

The Decision Tree (DT) (C4.5) consists of three parts: a decision node specifying test variables, a branch corresponding to one of the possible variable's values, and a leaf that contains the class of patient. At the start, the root of the tree is determined, then the node specified property is tested. The record of the patient is then classified in the same class based on the leaves probability [21]. The best attribute was used as a root for splitting. The objective criteria for judging the efficiency of the split is needed, and an information gain measure is used to select the test attribute at each node in the tree.

The Random Forest (RF) algorithm is one of the best among classification algorithms to classify large amounts of data with high accuracy. The basic principle is that a group of weak classifiers can be combined to form a strong classifier. It is an ensemble learning method building models that construct several decision trees at training time, and outputs the modal class out of the classes predicted by individual trees. It is a combination of tree predictors, where each tree depends on the values of a random vector sampled independently with the same distribution for all of the trees in the forest [22].

The Naïve Bayes (NB) classifier is a simple algorithm that is commonly used to predict the likelihood of group members. It is based on applying Bayes' theorem with a strong independent assumption. Bayes' theorem provides a way of calculating the conditional probability. In basic terms, a Naïve Bayes classifier assumes that the presence of a feature of a class is unrelated to the presence of any other feature. The Naïve Bayes classifier performs reasonably well even if the underlying assumption is not true [14].

The Bayesian Network (BNs) algorithm is a graphical structure that is used to represent knowledge, also known as belief networks (Bayes nets). It belongs to the family of probabilistic graphical models (GMs), with each node in the graph representing a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. BNs combines principles from graph theory, probability theory, statistics, and computer science [13].

The Multi-layer Perceptron (MLP) is a feedforward artificial neural network (ANN). It consists of several layers; each layer contains one or more nodes. The first is the input layer, the intermediates are the hidden layers, consisting of one or more layers, and finally the output layer contains the results of the computation. Each node connects with all other nodes in the subsequent layer, and it receives input from all other nodes in the previous layer, and so on [23].

2.4. The evaluation

The evaluation of the performance for any classification model depends on an external dataset that was not used for training. The evaluation indices are as follows: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The confusion metrics are derived from a basic data structure [2]. It contains information about actual and predicted class, calculated by a classification system. True Positive Rate (TPR) is referred to as Sensitivity, it is a probability that the result will be positive for varices. It is formulated by equation (3). The True Negative Rate (TNR) is the Specificity. It is the probability that a test result will be negative when the varices are not present. It is

[1] MedCalc Software is a developer of medical and statistical software solutions.

Table 1
Characteristics of the variables in the dataset.

Attributes (unit)	Mean ± Std. dev	Correlation Coefficient "r"	P-value
TSH	1.813 ± 1.140	0.001	0.6036
WBCx10^3mm^3	6.092 ± 5.564	−0.059	<0.0001
HbGL	13.490 ± 2.644	−0.055	<0.0001
Albumin (gdL)	3.79 ± 0.623	−0.280	<0.0001
Age (median, range)	54 (19–74)	0.034	0.0172
BMI	29.416 ± 4.268	−0.033	<0.0001
AST_ALT_Ratio	1.277 ± 0.693	0.055	0.0001
ALT Result	61.76 ± 39.078	−0.028	0.0052
AST Result	70.59 ± 43.695	−0.022	0.0109
Total Bilirubin (mgdl)	1.06 ± 0.608	0.194	<0.0001
AFP	18.923 ± 40.343	0.002	0.0041
Platelets (x10^3mm^3)	136.87 ± 64.69	−0.190	<0.0001
PC	80.27 ± 14.55	−0.190	<0.0001
Stiffness	27.32 ± 15.01	0.218	<0.0001
Baseline_PCR_log10	5.43 ± 0.94	−0.109	<0.0001
INR	1.203 ± 0.290	0.084	<0.0001
Creatine (mgdL)	0.870 ± 0.365	0.013	0.002
Glucose (mgdL)	104.75 ± 30.203	0.002	0.3794
Fib4Calculation	5.686 ± 25.668	0.043	<0.0001
ANA		0.043	0.0034
Negative	97.5%		
Positive	2.5%		
Gender		0.134	<0.0001
Female	31.3%		
Male	68.7%		
Liver		0.166	<0.0001
Abnormal	36.9%		
Cirrhotic	57.7%		
Normal	5.4%		
Diabetes		0.026	0.9408
No	78.3%		
Yes	21.7%		
Spleen		0.148	<0.0001
Average	36.6%		
Enlarged	61.7%		
Removed	1.7%		
Varices _group			
No varices	55.3%		
Varices	44.6%		
Total	100.0%		

expressed by equation (4); also

$$\text{Specificity} = (1 - \text{FPR}) \tag{2}$$

Precision is also referred to as Positive predictive value (PPV) and is expressed by equation (5). The accuracy is a probability of a total correct number of predictions, it is expressed as equation (6). The receiver operating characteristic (ROC) curve is a plot that displays the full picture of a trade-off between the sensitivity, true positive rate, (TPR) and (1- specificity), and false positive rate (FPR) across a series of cut-off points. FPR represents the x-axis and the TPR represents the y-axis. The ROC (Receiver Operating Characteristics) curve is plotted to find the probability of the prediction class to evaluate the performance of machine learning algorithms. The area under the ROC curve (AUC) [24] evaluates the ability to correctly detect diseased and non-diseased; it is considered an effective measure of the inherent validity of a diagnostic test. Also, it is used for finding an optimal cut-off point to least misclassify diseased and non-diseased. It illustrates the performance of a binary classifier system as its discrimination threshold is varied [24]. The optimal cut-offs were chosen when the sum of sensitivity and specificity would be maximum; PPVs and NPVs were computed for these values.

$$\text{Sensitivity (TPR)} = \frac{TP}{(TP + FN)} \tag{3}$$

$$\text{Specificity (TNR)} = \frac{TN}{(TN + FP)} \tag{4}$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{5}$$

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{6}$$

3. Results, analysis, and discussion

Four thousand, nine hundred and sixty-five patients were included in this study, with a median age of 54 years range (range 19–74). There were 31.3% female and 68.7% male. According to the ultrasonography, 36.6% of patients in the dataset had an average spleen while 61.7% had an enlarged spleen, and 1.7% had spleen removed. According to endoscopic findings, the patient's dataset contains two groups, varices and no varices. Characteristics of the attributes in the patient dataset after statistical analysis is shown in Table 1. The heatmap plot for Pearson's correlation coefficient between all variables is shown in Fig. 2. The Patient dataset was divided according to random sampling into two separated sets: 3972 patients and another 993 patients, about 80% for training and the rest for testing. The data were presented as mean value and standard deviation. The correlation and P-value results identified Gender, Platelet, Albumin, Total Bilirubin, Baseline_PCR, Liver, Spleen, Stiffness, and Prothrombin Concentration (PC) as independent predictors of varices. These attributes have a P-value less than 0.0001 and an accepted correlation coefficient of ($|r|>0.1$), as shown in Table 1.

The attributes selected by the feature selection approaches have been presented in Table 2, but the unselected attributes were deleted from this table. We compared the proposed filter method (P-value + CSF) as discussed in section 2.2 with other filter methods such as to measure Information Gain (IG) for each attribute with respect to the varices class, and principal component analysis PCA, IG and PCA used in conjunction with a Ranker search. Greedy Stepwise, PSO and GA were the search methods for the wrapper method; it evaluated attribute sets by using a learning algorithm (BN), 5-fold cross-validation to estimate the accuracy of the learning scheme for a set of attributes. The results of feature selection (filter and wrapper) methods are shown in Table 2. The best feature selection method should reach peak accuracy for the smallest number of attributes, so these selected attributes have been entered to the different ML algorithm as shown in Table 3.

It was noticed that (P-value + CSF) based on filter approach outperformed because it gave the best result with minimum number of features and independent of applied ML algorithm. It is particularity effective in computation time and robust to overfitting, works well with voluminous databases, is computationally less expensive, tends to give fast results on execution, is a scalable approach to large datasets, computationally efficient, works with low computational complexity, and is good generality. Hence, it was found that only nine attributes are more significant attributes.

For the first classification algorithm, the size of the Decision Tree (C4.5) was 408 and number of leaves were 219. After training and testing, the correctly classified instances were 667 or 67.17%, and incorrectly classified instances were 326 or 32.82%, where the values of the matrix were 401 for TN, 188 for FN, 138 for FP, and 266 for TP. For the second proposed classification model Naïve Bayes (NB), 662 (66.67%) and 331(33.33%) for the correctly and incorrectly classified instances, respectively, where values of the matrix were 411 for TN, 203 for FN, 128 for FP, and 251 for TP. For the Bayesian Network classification algorithm, after the training and testing process, the correctly classified instances were 684 (68.88%), and incorrectly classified instances were 309 (31.11%), where values of the matrix were 388 for TN, 158 for FN, 151 for FP, and 296 for TP. The Multilayer perceptron (MLP) classification algorithm was based on back-propagation neural networks. Its architecture has consisted of nine sigmoid nodes in one hidden layer, with 665 (66.97%) classified were correctly and 328 (33.03%) were incorrect, where values of the matrix were 369 for TN, 158 for FN,

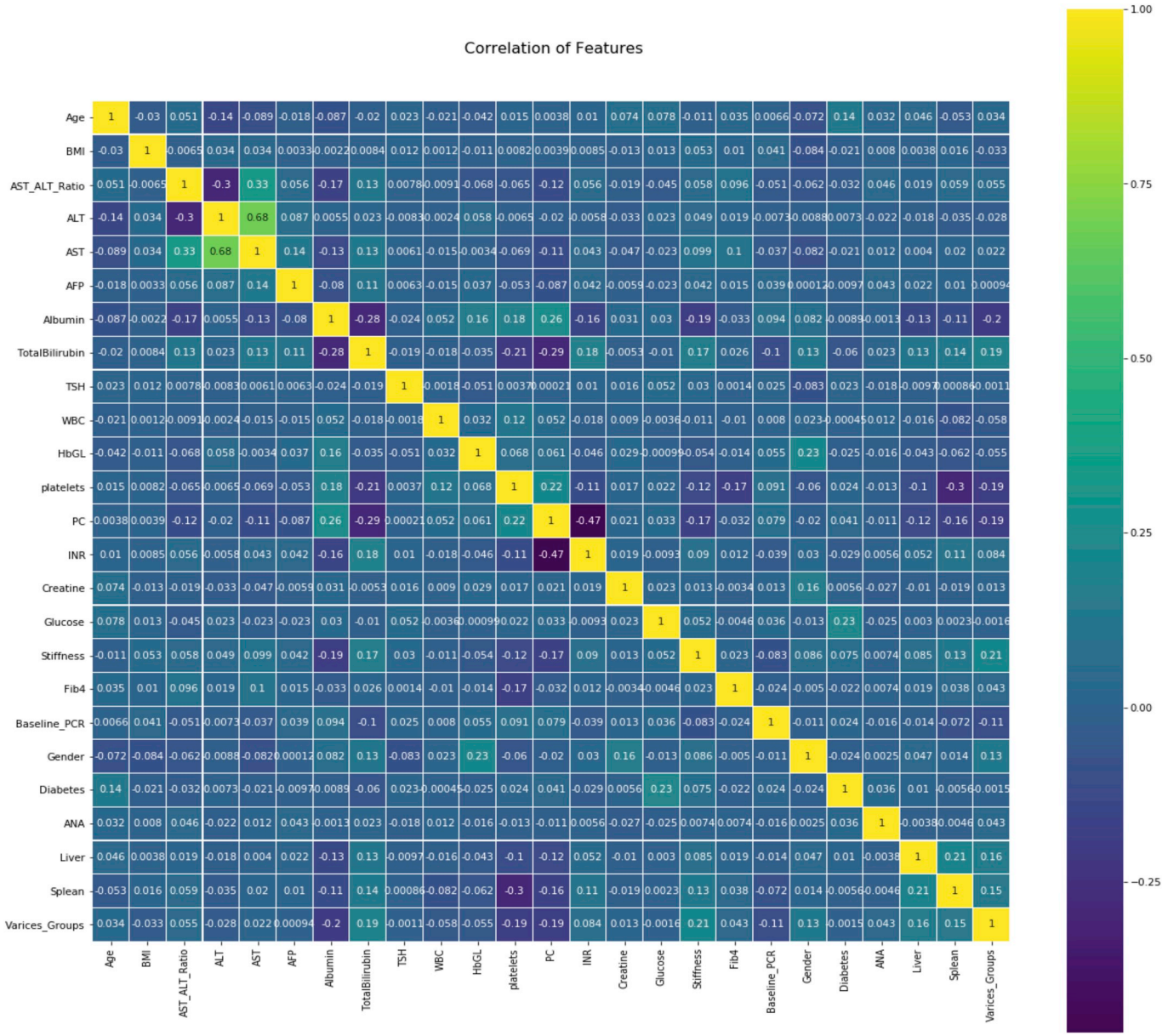

Fig. 2. Heatmap plot for correlation between all features.

Table 2
Attributes chosen by the different feature selection approaches.

Attributes	P-value + CSF	PCA	IG	GA	PSO	Stepwise
Gender	✓	✓	✓	✓	✓	✓
WBC			✓	✓		
Liver	✓	✓	✓	✓	✓	✓
Albumin	✓	✓	✓	✓	✓	✓
Baseline_PCR	✓	✓	✓		✓	
AST_ALT_Ratio		✓		✓	✓	
Fib4		✓	✓			
ANA		✓		✓	✓	
Total Bilirubin	✓	✓	✓	✓	✓	✓
INR		✓	✓			
Platelets	✓	✓	✓	✓	✓	✓
PC	✓	✓	✓		✓	
Stiffness	✓	✓	✓	✓	✓	✓
Spleen	✓	✓	✓	✓	✓	✓
HbGl				✓		✓
Ascites		✓			✓	✓
Total	9	14	12	11	12	9

170 for FP, and 296 for TP. Finally, when applying powerful algorithms like support vector machine (SVM) and Random Forest (RF), the result of sensitivity was 58.6, specificity 75.5, precision 67.7, AUC 67.1, accuracy 67.8 for the SVM algorithm Also, the result of RF was 58.4 for sensitivity, 72.9 for Specificity, 66.1 for precision, 71.1 for AUC, and 66.3 for accuracy.

Although there are various existing machine learning algorithms, these algorithms have their own limitations; numerous iterations, binning of the continuous arguments, and high computation time. We selected the six common algorithms to overcome various limitations including omission of complex iterative estimations of the parameter, that can be applied on a large dataset in real-time such as Naïve Bayes. In addition, these algorithms have many advantages such as being simple to understand, easy to implement, requiring little prior knowledge, able to handle both numerical and categorical data, robust, and able to address large and noisy datasets. Also, these common algorithms proved to have a wide success and gave better results with the medical dataset for diagnosis of disease. Our proposed classification models were evaluated by the accuracy and other factors as shown in Table 4. Sensitivity,

Table 3
Performance proposed algorithms with different types of feature selection.

	P-value + CSF (9)	PCA (14)	IG (12)	GA (11)	PSO (12)	Stepwise (9)
SVM						
AUC	67.1	66.3	67.0	65.9	66.4	66.0
Acc.	67.8	67.1	67.7	66.7	67.2	66.9
RF						
AUC	71.1	71.9	71.1	70.4	71.2	70.3
Acc.	66.3	65.6	65.8	65.8	65.9	65.4
C4.5						
AUC	69.0	68.4	63.6	68.6	69.3	86.8
Acc.	67.2	65.2	61.7	67.0	66.2	67.4
MLP						
AUC	71.6	71.9	71.0	72.9	70.1	72.5
Acc.	65.6	67.8	66.6	68.4	66.2	66.9
NB						
AUC	73.2	72.7	72.3	71.7	72.9	72.4
Acc.	66.7	65.7	66.5	66.6	66.0	67.3
BNs						
AUC	74.8	74.5	74.5	74.0	74.7	74.2
Acc.	68.9	67.5	67.8	68.0	68.4	68.2

Table 4
Performance of the Classification algorithms.

ML Algorithm	Sensitivity	Specificity	Precision	AUC	Accuracy
SVM	58.6	75.5	67.7	67.1	67.8
RF	58.4	72.9	66.1	71.1	66.3
C4.5	58.6	74.4	67.1	69.0	67.2
MLP	67.4	64.0	66.0	71.6	65.6
NB	55.3	76.3	66.7	73.2	66.7
BNs	65.3	72.0	68.8	74.8	68.9

specificity, precision, recall, Area Under ROC (AUC) analysis, and the accuracy of the models have been calculated for predicting presence of varices. The results were similar, but the Bayesian Net algorithm classifiers were better efficiently and effectively than the others. A comparison between the ROC curves for the best four classifiers is shown as Fig. 3. The area under ROC curve for Bayesian Net of 74.8% was the best model. AUC curves of The Multilayer perceptron, Naïve Bayes and Bayesian Net were close, but the ROC of decision tree algorithm was lowest in value. The Bayesian Nets algorithm performed to exhibit better performance than the other classifiers. It achieved the highest accuracy of 68.9% and 74.8% for (AUC).

In order to obtain more accurate results and use all available records of data, 10-fold cross-validation was utilized to avoid the overtraining problem. We applied the BN_S algorithm on a dataset with the nine variables using 10-fold cross-validation. It achieved 72.8% AUC and 68% accuracy, which is very close to the results of using training and test sets separately. To check from our predictive model what is best suitable for predicting the esophageal varices, four additional predictive models were selected based on a literature review. ROC curve analysis of the four models was compared with our model. Our proposed model yielded higher AUC from the predictive models of Al Ghamdi et al. [25], Sedrak et al. [26], and Bledar et al. [11], while our AUC was less than that of Wang et al. [27]. The result of Wang et al. may have been inaccurate because the dataset was small and it was from a single hospital and didn't cover a wide area, unlike our study. Our dataset was collected from approximately fifteen hospitals around Egypt. Even though there

Table 5
Comparing the results of this paper with other predictive methods.

Model [Ref.]	Significant features	Dataset	Used Algorithm	AUC (%)
Al Ghamdi et al. (2017) [25]	Stiffness	75	Filter approach	67
Bledar et al. (2017) [11]	FIB4	139	Binary logistic regression	66
Sedrak et al. (2015) [26]	AST to Platelet ratio index (APRI)	74	logistic regression analysis	61
Wang et al. (2014) [27]	Portal vein diameter, hemoglobin level, PLt. count/Spl. diameter	104	multivariate logistic regression analysis	81.4
Our study	Plt., Albu., T Biliru., PC, Spl., B_PCR., Liver, Stiff., and Gen.	4962	BNs	74.8

Plt. Platelet, Albu. Albumin, T Biliru. Total Bilirubin, B_PCR. Baseline_PCR, Spl. Spleen, Stiff. Stiffness, Gen. Gender, AST Aspartate Aminotransferase.

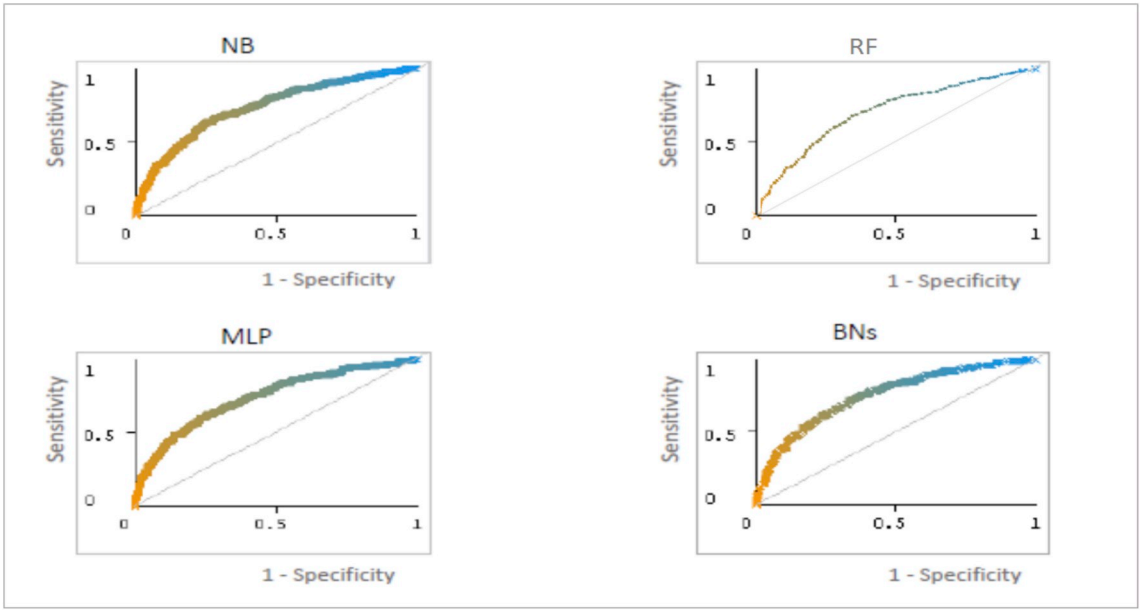

Fig. 3. ROC curves plots for our algorithms; Naïve Bayes (NB) = 73.2, Random Forest (RF) = 71.1, Multilayer perceptron (MLP) = 71.6, and Bayesian Net (BNs) = 74.8.

was no significant difference from the four predictive models concerning the used approach, our model achieves the highest AUC value as shown Table 5, thus representing the best predictive model for esophageal varices in liver cirrhosis patients. Among the tested five predictive models, as well as other individual and compound variables, our proposed predictive model had the greatest diagnostic ability in predicting esophageal varices.

4. Conclusion

In this paper, six feature selection methods were proposed; three of them depend on the filter and the other three depend on the wrapper approach. In the filter approach, we used (p-value) +CFS, Information Gain (IG) and Principal Components Analysis (PCA). In the wrapper approach, we used a greedy stepwise, Genetic algorithm (GA), and Particle Swarm Optimization (PSO). Six machine learning algorithms were utilized for predicting presence of Esophageal Varices in chronic hepatitis C patients, as alternative methods with respect to diagnostics with the upper gastrointestinal endoscopy. These algorithms consisted of the multilayer perceptron, support vector machine, random forest, Decision tree, Naïve Bayes and Bayesian Nets. A comparison between six algorithms with six feature selection methods has been carried out for predicting varices. The parameters Gender, Platelet, Albumin, Total Bilirubin, Baseline_PCR, Liver, Spleen, Stiffness, and prothrombin concentration were found to be the most significant independent predictors for varices. It is clear that the Bayesian Network is more accurate as compared to use of each algorithm alone. The Bayesian Net algorithm achieved 68.9% for accuracy and 74.8% for the AUC. To avoid the risks of gastrointestinal screening, the proposed models could be used as an acceptable, low cost, and safe method for predicting Esophageal varices in cirrhotic patients. Non-invasive markers may be useful as a first step to identify varices for cirrhotic patients and to reduce the number of upper endoscopies. However, we are still far from the possibility of replacing upper endoscopy with non-invasive markers in most patients.

Ethical statement

There is no extra ethical statement to make.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

No conflict of interest.

Acknowledgement

We acknowledge the Egyptian National Committee for Control of Viral Hepatitis for supplying the clinical data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2019.100267.

References

[1] de Franchis R. Baveno VI Faculty. Expanding consensus in portal hypertension: report of the Baveno VI Consensus Workshop: stratifying risk and individualizing care for portal hypertension. J Hepatol 2015;63:743–52. https://doi.org/10.1016/j.jhep.2015.05.022.

[2] Peter Drotar ZS. Comparitive study of machine learing technique for supervised classification. Acta Electrotech Inf 2014;14:5–10. https://doi.org/10.15546/aeei-2014-0021.

[3] ElHefnawi M, Abdalla M, Ahmed S, Elakel W, Esmat G, Elraziky M, et al. Accurate prediction of response to interferon-based therapy in Egyptian patients with chronic hepatitis C using machine-learning approaches. IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM 2012. IEEE Comput Soc 2012:803–10. https://doi.org/10.1109/ASONAM.2012.140.

[4] El Houby EMF. A framework for prediction of response to hcv therapy using different data mining techniques. Adv Bioinform 2014;2014:10. https://doi.org/10.1155/2014/181056.

[5] Hashem S, Esmat G, Elakel W, Habashy S, Raouf SA, ElHefnawi M, et al. Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis C patients. IEEE ACM Trans Comput Biol Bioinform 2018;15:861–8. https://doi.org/10.1109/TCBB.2017.2690848.

[6] Soliman OS, Elhamd A. Classification of hepatitis C virus using modified Particle Swarm optimization and least squares support vector machine. Int J Sci Eng Res 2014;5:122–9.

[7] Hashem S, Esmat G, Elakel W, Habashy S, Abdel Raouf S, Darweesh S, et al. Accurate prediction of advanced liver fibrosis using the decision tree learning algorithm in chronic hepatitis C Egyptian patients. Gastroenterol Res Pract 2016: 2016. https://doi.org/10.1155/2016/2636390.

[8] Khairy M, Abdel-Rahman M, El-Raziky M, El-Akel W, Zayed N, Khatab H, et al. Non-invasive prediction of hepatic fibrosis in patients with chronic HCV based on the routine pre-treatment workup. Hepat Mon 2012;12:0–5. https://doi.org/10.5812/hepatmon.6718.

[9] Stefanescu H, Grigorescu M, Lupsor M, Maniu A, Crisan D, Procopet B, et al. A new and simple algorithm for the noninvasive assessment of esophageal varices in cirrhotic patients using serum fibrosis markers and transient elastography. J Gastrointest Liver Dis: JGLD 2011;20:57–64.

[10] Sebastiani G, Tempesta D, Fattovich G, Castera L, Halfon P, Bourliere M, et al. Prediction of oesophageal varices in hepatic cirrhosis by simple serum non-invasive markers: results of a multicenter, large-scale study. J Hepatol 2010;53:630–8. https://doi.org/10.1016/j.jhep.2010.04.019.

[11] Kraja Bledar, Mone Iris, Akshija Ilir, Koçollari Adea, Skerdi Prifti GB, Bledar. Predictors of esophageal varices and first variceal bleeding in liver cirrhosis patients. World J Gastroenterol 2017:23. https://doi.org/10.3748/wjg.v23.i26.4806.

[12] Xu XD, Dai JJ, Qian JQ, Pin X, Wang WJ. New index to predict esophageal variceal bleeding in cirrhotic patients. World J Gastroenterol 2014;20:6989–94. https://doi.org/10.3748/wjg.v20.i22.6989.

[13] Banu Priya M, Laura Juliet P, Tamilselvi PR. Performance analysis of liver disease prediction using machine learning algorithms. Int. Res. J. Eng. Technol. 2018;5:206–11.

[14] Vijayarani S, Dhayanand S. Liver disease prediction using SVM and Naïve Bayes algorithms. Int. J. Sci. Eng. Technol. Res. 2015;4:816–20.

[15] Fatima M, Pasha M. Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl 2017;09:1–16. https://doi.org/10.4236/jilsa.2017.91001.

[16] Kiran Kumar M, Sreedevi M, Reddy YCAP. Survey on machine learning algorithms for liver disease diagnosis and prediction. Int J Eng Technol 2018;7:99–102. https://doi.org/10.1016/j.neuroimage.2012.03.068.

[17] Jain D, Singh V. Feature selection and classification systems for chronic disease prediction: a review. Egypt Inf J 2018;19:179–89. https://doi.org/10.1016/j.eij.2018.03.002.

[18] Software M. Software manual MedCalc 2017. Belgium: 1993-2017 MedCalc software bvba all. 2017.

[19] Lind Douglas A, William G. SAW. Statistical techniques in business and economics. Fifteenth. Americas, New York,: McGraw-Hill/Irwin, a bTim vertovec. 1221Avenue of the: usiness unit of The McGraw-Hill Companies, Inc; 2012.

[20] Harb HM, Desuky AS. Feature selection on classification of medical datasets based on Particle Swarm optimization. Int J Comput Appl 2014;104:975–8887. https://doi.org/10.5120/18197-9118.

[21] Venkata Ramana B, Babu Ms Prasad, Venkateswarlu NB, Professor A. A critical study of selected classification algorithms for liver disease diagnosis. Int J Database Manag Syst 2011:3. https://doi.org/10.5121/ijdms.2011.3207.

[22] Jain A. Machine learning techniques for medical diagnosis: a review. In: International Conference of science,technology and management (ICSTM), Center New Delhi, India; 2015. p. 2449–59.

[23] Belciug S, Gorunescu F. Bayesian-driven multi-layer perceptron applied to liver fibrosis stadialization, vol. 38; 2014.

[24] Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr 2011;48:277–87.

[25] Hassan Al Ghamdi M, Wang L, Hu JW, Dong S, Jian YC, Hu L, et al. Transient elastography (fibroscan) compared to diagnostic endoscopy in the diagnosis of varices in patients with cirrhosis. Sci J Clin Med 2017;5:55–9. https://doi.org/10.11648/j.sjcm.20160506.13.

[26] Sedrak H, Khalifa R, Elkafrawy A, Elewa H. Noninvasive predictors of large esophageal varices: is there an emerging role of aspartate aminotransferase-to-platelet ratio index in hepatocellular carcinoma? Egypt J Intern Med 2015;27:139. https://doi.org/10.4103/1110-7782.174935.

[27] Wang L, Hu J, Dong S, Jian YC, Hu L, Yang G, et al. Noninvasive prediction of large esophageal varices in liver cirrhosis patients. Clin Investigative Med Med Clin Exp 2014;37:E38–46. https://doi.org/10.4103/1319-3767.58767.