Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1351863

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 7 - Issue 3: 445-456 / May 2024

USAGE OF WEKA SOFTWARE BASED ON MACHINE LEARNING ALGORITHMS FOR PREDICTION OF LIVER FIBROSIS/CIRRHOSIS

Rukiye UZUN ARSLAN[1]*, Ziynet PAMUK[2], Ceren KAYA[2]

[1]Zonguldak Bulent Ecevit University, Faculty of Engineering, Department of Electrical and Electronics Engineering, 67100, Zonguldak, Türkiye
[2]Zonguldak Bulent Ecevit University, Faculty of Engineering, Department of Biomedical Engineering, 67100, Zonguldak, Türkiye

Abstract: The liver, a life-sustaining organ, plays a substantial role in many body functions. Liver diseases have become an important world health problem in terms of prevalence, incidences, and mortalities. Liver fibrosis/cirrhosis is great of importance, because if not treated in time liver cancer could be occurred and spread to other parts of the body. For this reason, early diagnosis of liver fibrosis/cirrhosis gives significance. Accordingly, this study investigated the performances of different machine learning algorithms for prediction of liver fibrosis/cirrhosis based on demographic and blood values. In this context, random forest, k-nearest neighbour, C4.5 decision tree, K-star, random tree and reduced error pruning tree algorithms were used. Two distinct approaches were employed to evaluate the performances of machine learning algorithms. In the first approach, the entire features of dataset were utilized, while in the second approach, only the features selected through principal component analysis were used. Each approach was rigorously assessed using both 10-fold cross-validation and data splitting (70% train and 30% test) techniques. By conducting separate evaluations for each approach, a comprehensive understanding of the effectiveness of utilizing all features versus extracted features based principal component analysis was attained, providing valuable insights into the impact of feature dimensionality reduction on model performance. In this study, all analyses were implemented on WEKA data mining tool. In the first approach, the classification accuracies of random forest algorithm were 89.72% and 90.75% with the application of data splitting (70%-30%) and cross-validation techniques, respectively. In the second approach, where feature reduction is performed using principal component analysis technique, the accuracy values obtained from data splitting and cross-validation techniques of random forest algorithm were 88.61% and 88.83%, respectively. The obtained results revealed out that random forest algorithm outperformed for both approaches. Besides, the application of principal component analysis technique negatively affected the classification performance of used machine learning algorithms. It is thought that the proposed model will guide specialist physicians in making appropriate treatment decisions for patients with liver fibrosis/cirrhosis, potentially leading to death in its advanced stages.

Keywords: WEKA, Liver fibrosis/cirrhosis, Principal component analysis, Feature selection, Early diagnosis, Machine learning

***Corresponding author:** Zonguldak Bulent Ecevit University, Faculty of Engineering, Department of Electrical and Electronics Engineering, 67100, Zonguldak, Türkiye
E mail: rukiyeuzun67@gmail.com (R. UZUN ARSLAN)
Rukiye UZUN ARSLAN https://orcid.org/0000-0002-2082-8695
Ziynet PAMUK https://orcid.org/0000-0003-3792-2183
Ceren KAYA https://orcid.org/0000-0002-1970-2833

Received: August 29, 2023
Accepted: April 02, 2024
Published: May 15, 2024

Cite as: Uzun Arslan R, Pamuk Z, Kaya C. 2024. Usage of weka software based on machine learning algorithms for prediction of liver fibrosis/cirrhosis. BSJ Eng Sci, 7(3): 445-456.

1. Introduction

The liver is a vital organ placed in the upper right side of the abdomen, beneath the diaphragm. It is the largest internal organ in the human body, and its weight is about 1.5 kg, although it differs in men and women. The liver, which can both expand and renew itself, undertakes many important vital functions such as clearing toxic wastes in the blood, storing some important vitamins, storing, and digesting fats in the body, making bile acids usable for our body, balancing hormones. If the liver cannot perform these functions properly, it leads several types of liver diseases like fatty liver, fibrosis, cirrhosis, and liver cancer (Lin, 2009; Acarlı, 2020). These diseases have become a significant global health concern regarding their prevalence, incidences, and mortalities (Asrani et al., 2019; Del Campo et al., 2018; Ramana et al., 2011). The progression of liver disease is given in Figure 1.

Fibrosis and cirrhosis are conditions involved in scarring of the liver tissue. As seen in Figure 1, since fibrosis/cirrhosis is the previous stage of liver cancer, early diagnosis is of great importance as it can increase the survival rate. Nowadays, liver biopsy accepted as the gold standard test has been used. However, biopsy has many disadvantages such as the risk of complications, being invasive, not being repeated frequently, not providing information about the whole liver, and high cost. Therefore, it is not possible to establish a final diagnosis by liver biopsy. To overcome these disadvantages, alternative methods (biochemical test, biomedical imaging techniques etc.) have been developed. In traditional biomedical tests, many clinical parameters such as alkaline phosphatase (ALP), Alanine amino-transferase (ALT), Aspartate amino-transferase (AST) are extensively utilized in the diagnosis of the disease and in research studies. On the other hand, the developments in technology have made popular the usage of machine learning (ML), artificial intelligence (AI) and data mining techniques, especially in the field of medicine to diagnose and treat disease. Thanks to these techniques, disease management can be carried out quickly and accurately by a specialist physician. On the other hand, different signs, symptoms, and enzyme levels can be encountered in the diagnosis of liver diseases (Schiff et al., 2017). This situation causes the disease process to be error-prone and complex. Because of this, utilizing ML and data mining algorithms can give support to specialist physicians in diagnosis and prediction of liver disease. In this context, the performances of six different ML algorithms in classifying healthy and unhealthy (liver fibrosis/cirrhosis) patients were investigated within the scope of this study. This classification study performed in WEKA software was carried out using random forest (RF), C4.5 decision tree (J48), random tree (RT), K-star, reduced error pruning tree (REPTree) and k-nearest neighbour (k-NN, IBk) algorithms. The dataset of 1200 patient collected from Zonguldak Ataturk State Hospital, Türkiye was used in the study, in which all necessary ethics committee permissions were obtained. The dataset consisted of features including age, gender, aspartate aminotransferase (AST), alanine aminotransferase (ALT), Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase (ALP) and Albumin.

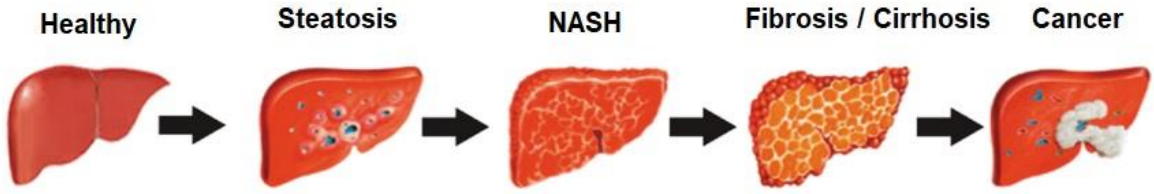


Figure 1. Progression of liver diseases (Xie et. al, 2016).

The remaining parts of this study were formed as follows: In the second part, studies on the prediction of liver diseases used ML methods were given. In the third part, brief information was given about the data set utilized in the study, which pre-processing it has been through, the models created, and the algorithms used in the model. In the fourth part, the performance metrics, complexity matrices and graphs obtained from the algorithms were interpreted. In the last part, the obtained were discussed and future studies were mentioned.

2. Related Works

In literature, many studies-based ML and data mining have been realized to predict liver diseases. Gulia et al. (2014) proposed a hybrid model with (having) three stages (classification, feature selection and comparison of results) to detect potential liver patient using WEKA software. They showed that random forest gives better results (accuracy with 71.87%) than the other methods. Alkuşak and Gök (2014) studied liver failure detection utilizing ML methods on WEKA. Using two different datasets, they evaluated performances of the methods and found that the neural networks' performances were high for both datasets. Pahareeya et al. (2014) carried out four different ML algorithms on Indian Liver Patient Dataset (ILPD) using 10-fold cross-validation technique. They reached 89.11% accuracy value with RF algorithm in liver disease classification. Borulday et al. (2017) investigated the decision tree algorithm's performance to detect different liver diseases such as cirrhosis and acute hepatitis. In addition, an interface has been developed that includes different demographic data, physical examination, laboratory tests, symptoms, radiological imaging, liver biopsy and unclassified data. Thus, specialist doctors were given the opportunity to select tabs in the interface. Muthuselvan et al. (2018) investigated binary classification performances of Naïve Bayes (NB), K-star, J48 and Random Tree (RT) algorithms by using WEKA software on ILPD dataset. The obtained results demonstrated that RT had superior accuracy (74.2%) with nominal execution time. Alaybeyoğlu and Mulayim (2018) developed a smart model for liver cancer diagnosis based on support vector machine algorithm and reached an accuracy of 78.33%. Ma et al. (2018) utilized diverse ML algorithms to obtain a useful monitoring and optimal predictive model of non-alcoholic fatty liver disease. The algorithms were carried out through WEKA, the classification was implemented by using a 10-fold cross-validation technique and the performances of algorithms were compared. They revealed out that the Bayesian network model (by 83% accuracy) has best performance as compared the other algorithms. Rahman et al. (2019) evaluated the performance of many different ML algorithms to estimate and reduce the cost of diagnosing chronic liver disease. ILPD was used as the data set and the highest performance was obtained in logistic regression (75% accuracy). Keleş et al. (2020) investigated various ML algorithms' performances (like J48, RF, RT) implemented in WEKA to diagnose liver diseases. Using ILPD, the performances of the algorithms were compared regarding the evaluation metrics. It was obtained from the analysis that the algorithm with the highest performance was RF with 81.9% accuracy. Gaber et al. (2022) proposed a computer-aided decision support model based on ML algorithms and voting classifier to detect fatty and normal livers from ultrasound liver images. The classification performances were obtained 95.71% and 93.12% in voting classifier and J48 algorithm, respectively. Velu et al. (2022) presented a novel classification model on the basis of liver function test results to detect probable liver patients. The model showed a success rate of 98.4% and 99.36% for the test and training phases, respectively. They also designed an interface providing the opportunity to enter patient information for specialist. Dritsas and Trigka (2023) examined the performance of different ML algorithms in the prediction of liver disease at an early stage. The algorithms performances were evaluated regarding accuracy, precision, recall, F-measure and area under the curve (AUC) by 10-fold cross-validation. They revealed out that voting algorithm has higher performance with 80.1% of accuracy in comparison to other algorithms. Mukhyber et al. (2023) investigated the performance of several data mining algorithms in the early diagnosis of liver disease according to parameters such as accuracy, precision, recall and f-measure. In addition, they addressed the comparison of algorithms' performances in accordance with the performance of ILPD.

The above-mentioned studies clearly show that ML algorithms have been widely used in the diagnosis of liver diseases. The models having superior performance had been determined according to different metrics such as accuracy, precision, specificity and F-measure. In this context, accuracy, precision, recall, F-measure and receiver operating characteristic curve (ROC) metrics were used within the scope of this study. Besides, it has been determined that the obtained results in the above-mentioned studies varies depending on many factors like the data set used, evaluation metrics, and the type of liver disease examined.

3. Materials and Methods

This study analyses the individuals' health condition (healthy and unhealthy-suffering from liver fibrosis/cirrhosis) by using 6 different ML algorithms (RF, IBk, k-Star, RT, REPTree and J48) based two different approaches. In the first approach all the features in the dataset were used, while the selected features extracted by PCA were utilized in the second approach. The performances of ML algorithms for each approach were evaluated regarding cross-validation (10-fold) and data splitting (70% train-30% test) techniques separately. General flowchart of proposed model was shown in Figure 2.

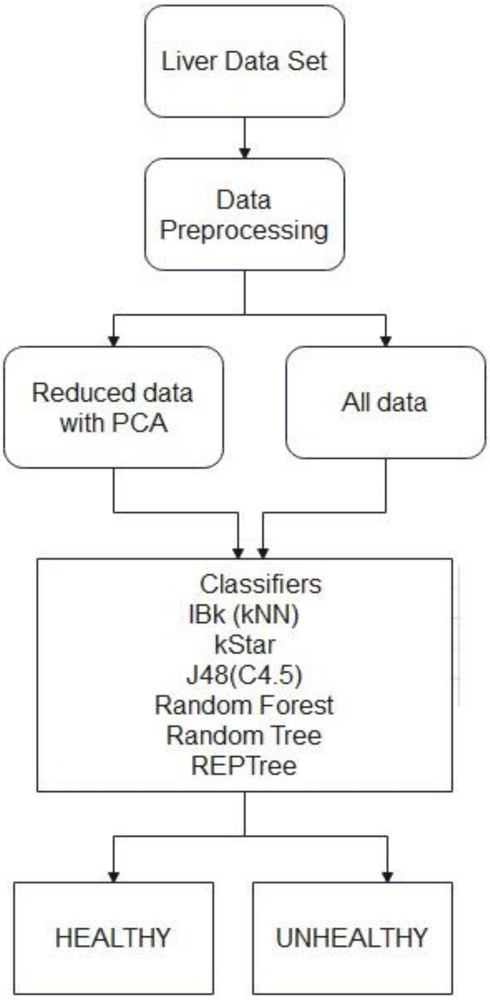


Figure 2. General flowchart of proposed model

3.1. Data Set

This study carried out by using a balanced dataset (600 healthy, 600 with Liver Fibrosis/Cirrhosis) collected from Zonguldak Ataturk State Hospital. The required ethical approval for performing this study was obtained from Zonguldak Bulent Ecevit University, Non-interventional Clinical Research Ethics Committee (Decision no: 2021/05; Date: 10.03.2021). Age, gender, aspartate aminotransferase (AST), alanine aminotransferase (ALT), Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase (ALP) and Albumin features were used to comprise the required dataset. The reference values of these features, which were determined based on the experience of specialist physicians in accordance with ICD-10 K74 standards, were presented in Table 1 with data types.

3.2. Feature Selection

Feature selection indeed plays a crucial role in various machine learning tasks, including classification (Ucar and İncetaş, 2022). By identifying and utilizing only the most relevant features, one can improve model performance, reduce computational complexity, and potentially enhance interpretability (Bulut et. al, 2023).

On the other hand, principal component analysis (PCA) is a technique used to increase interpretability by reducing the size of datasets and at the same time minimizing information loss. It minimizes information loss by creating new uncorrelated variables that successively maximize variance. When feature selection and PCA are used together, better model performance and more effective data analysis can be achieved. In this context, in this study, a dataset with eight features was reduced to six features by PCA.

Table 1. The reference values and data types of features used in dataset

Features	Age	Gender	Reference Value	Data type
Age	-	-	-	Integer
Gender	-	-	-	{ 1, 0 }(Female, Male)
AST	Adult	M	0-35 U/L	Real
		F	0-31 U/L	
ALT	Adult	M	0-45 U/L	Real
		F	0-34 U/L	
T.Bilirubin	-	M/F	0-0.2 mg/dL	Real
D.Bilirubin	Adult	M/F	0.3-1.2 mg/dL	Real
ALP	Adult	M/F	30-120 U/L	Integer
Albumin	Adult	M/F	3.5-5.2 g/dL	Real
	>60		3.2-4.6 g/dL	

3.3. Classification Algorithms

In this study, it was investigated which ML algorithm shows higher performance when making binary classification of patients (healthy and unhealthy-suffering from liver fibrosis/cirrhosis). To that end, RF, IBk, k-Star, RT, REPTree and J48 algorithms were used considering the previous studies in the literature. These algorithms are briefly described below.

RT, an ensemble learning algorithm, generates rules by learning multiple individual rules. It employs a bagging idea to create a decision tree and produce a random dataset. While in a standard decision tree, each node is split using the best split among all variables, in RT algorithm, each node is split using the best split among a randomly selected subset of variables at that node. This approach enables RT to achieve high accuracy results (Işık and Ulusoy, 2021).

RF algorithm is a supervised ensemble learning method developed by Breiman and used to solve classification and/or regression tasks (Breiman, 2001). RF, having common usage, consists of many decision trees. The biggest disadvantage of the decision tree algorithm, one of the classical machine learning methods, is memorizing the data due to excessive learning. In order to overcome this disadvantage, the RO algorithm divides both the data set and its features into a large number of random subsets, trains them, and arrives at a general decision by averaging the answers of each subset generated by the decision tree (Sevli, 2019).

k-NN algorithm is a supervised learning algorithm used for both classification and regression. The working principle of this algorithm is relied on that data points with similar features tend to belong to the same class or have similar output values. Given a new data point, the algorithm finds the k closest data points in the training dataset based on a distance metric and then assigns the class label (in the case of classification) or predicts the value (in the case of regression) based on the majority vote or average of the labels/values of the k nearest neighbours (Duda et al., 2000; Uzun et. al, 2018; Azam et al., 2020).

K-star algorithm, a classification method in machine learning, relies on the k-nearest neighbor (k-NN) principle but distinguishes itself by employing the Pearson correlation coefficient as the similarity measure rather than traditional distance metrics. It assesses similarities between feature vectors focusing solely on linear relationships, generating a correlation matrix to measure interactions among instances in the training dataset. Utilizing these similarities, it classifies test instances by comparing them with the matrix. K-star can be particularly effective for small-sized datasets and situations where nonlinear relationships are deemed insignificant for classification (Ünal et. al, 2019; Mishra et. al, 2021).

J48, which is an implementation of the C4.5 algorithm, is a popular decision tree algorithm used for classification. It builds a decision tree by applying a recursive binary splitting approach. It uses an attribute selection measure, typically based on information gain or gain ratio, to determine the most informative attribute to divide the data at each node of the tree. The dividing process continues recursively until a stop criterion is met, such as reaching a maximum depth or a minimum number of instances per leaf (Kaya et al., 2017; Muthuselvan et al., 2018).

REPTree is a decision tree learning algorithm designed to minimize variance during tree construction. It operates by iteratively generating multiple decision trees and selecting the best one based on variance reduction. The algorithm employs the square of the average error in predictions as the criterion for pruning the tree, ensuring that the selected tree maintains high accuracy. It builds the decision tree based on information gain as the splitting criterion and utilizes reduced error pruning to streamline the learning process. By employing this approach, REPTree aims to produce decision trees that are efficient, accurate, and resistant to overfitting, making it a valuable tool in various machine learning applications, particularly in classification tasks (Şenel et. al, 2021).

All these algorithms, implemented on WEKA software, were used for classification of patients as healthy and unhealthy (suffering liver fibrosis/cirrhosis). WEKA is a popular open-source software tool. It comprises many machine learning algorithms, data pre-processing tools, and visualization abilities (Uzun et. al, 2019).

3.4. Evaluation Metrics

The ML algorithms' performances were evaluated by considering the accuracy, specificity, precision, recall and F-measure metrics that are frequently used in the similar studies. Accuracy (equation 1) is the ratio of correctly classified instances out of the total number of instances, whereas precision (equation 2) is the ratio of correctly predicted positive instances to all positive instances. Recall (equation 3) is the rate of the correctly predicted positive instances to all actual positive instances and F-measure (equation 4) is the harmonic mean of sensitivity and precision metrics. The mathematical equations for these metrics were given below. The confusion matrix evaluated the performance of ML algorithms consists of four elements: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). In the confusion matrix, TP, FN, FP and TN correspond to correctly predicted positive class value, incorrectly predicted negative class value, incorrectly predicted positive class value and correctly predicted negative class value, respectively (Narin et al., 2021; Şenyer Yapıcı, 2021).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

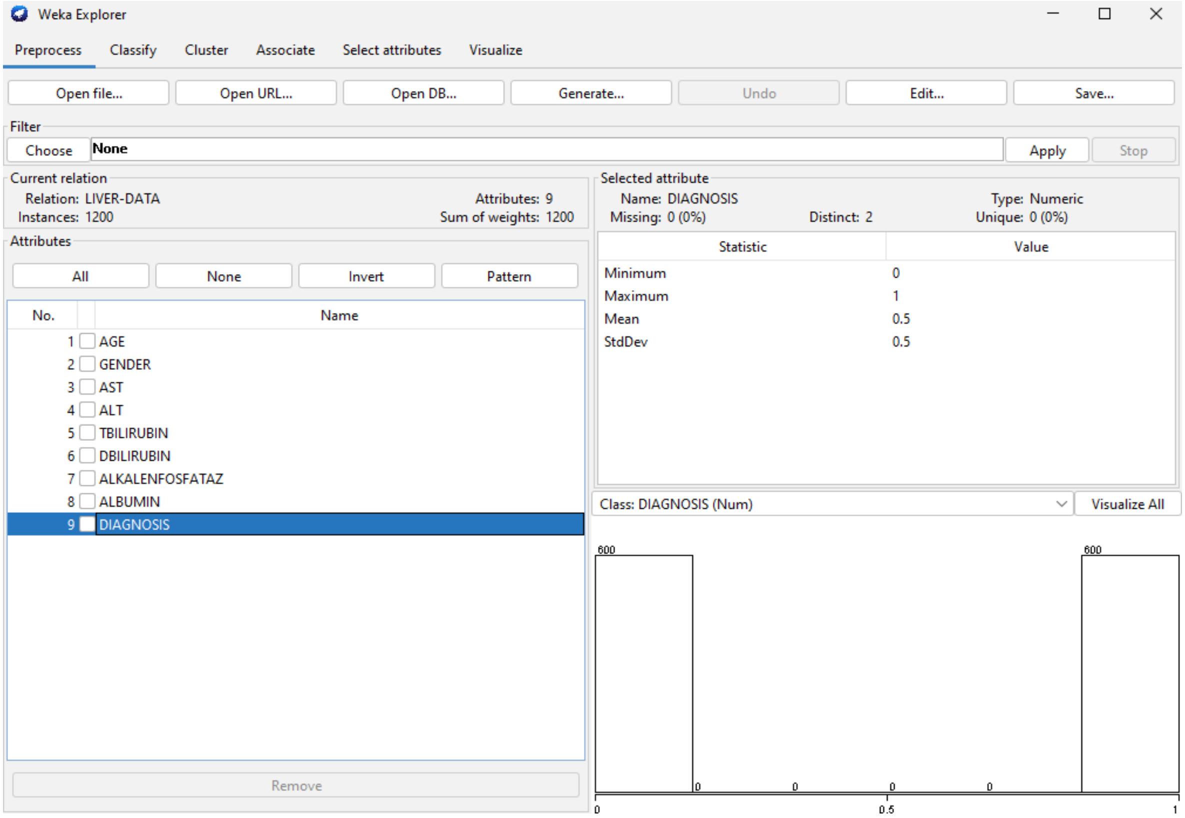
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{FMeasure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

In this study, the performances of ML algorithms were evaluated in accordance with cross-validation (10-fold) and data splitting (70%-30%) techniques. Besides, ROC metric with a value ranging from [0, 1] was given for the ML algorithm with the highest classification performance.

4. Results

In this study, ML algorithms performances was evaluated on a balanced dataset including liver fibrosis/cirrhosis patients (unhealthy, 600) and healthy patients (600) regarding cross-validation (10-fold) and data splitting (70%-30%) techniques separately. The classification performances of these algorithms were calculated based evaluation metrics given above. The dataset consisted of features including age, gender, aspartate aminotransferase (AST), alanine aminotransferase (ALT), Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase (ALP) and Albumin. These eight features were used as input parameters in WEKA (Figure 3).



Filter: None

Current relation
Relation: LIVER-DATA
Attributes: 9
Instances: 1200
Sum of weights: 1200

Attributes

No.	Name
1	☐ AGE
2	☐ GENDER
3	☐ AST
4	☐ ALT
5	☐ TBILIRUBIN
6	☐ DBILIRUBIN
7	☐ ALKALENFOSFATAZ
8	☐ ALBUMIN
9	☐ DIAGNOSIS

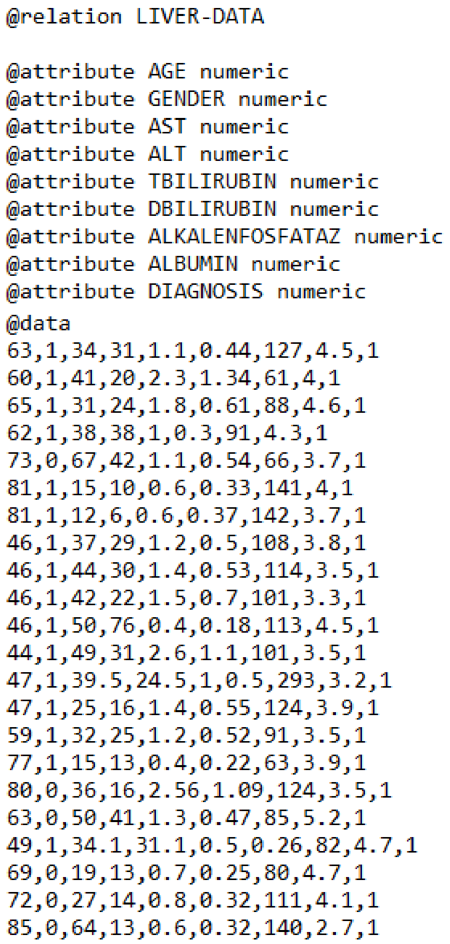
Selected attribute
Name: DIAGNOSIS
Type: Numeric
Missing: 0 (0%)
Distinct: 2
Unique: 0 (0%)

Statistic	Value
Minimum	0
Maximum	1
Mean	0.5
StdDev	0.5

Class: DIAGNOSIS (Num)

Figure 3. WEKA screenshot with all features given as input.

An example ARFF file of these features was shown in Figure 4. PCA method was used while feature selection. Ranker Search Method was selected when performing this process. The most significant feature was identified as Direct Bilirubin (0.64814) considering rank sorting of all features. The screenshot of ranked features appearing in WEKA after this process was illustrated in Figure 5. As clearly seen in Figure 5, age and ALT features, which their ranks were below 0.1, were not included in the reduced data set (six features).



```
@relation LIVER-DATA

@attribute AGE numeric
@attribute GENDER numeric
@attribute AST numeric
@attribute ALT numeric
@attribute TBILIRUBIN numeric
@attribute DBILIRUBIN numeric
@attribute ALKALENFOSFATAZ numeric
@attribute ALBUMIN numeric
@attribute DIAGNOSIS numeric
@data
63,1,34,31,1.1,0.44,127,4.5,1
60,1,41,20,2.3,1.34,61,4,1
65,1,31,24,1.8,0.61,88,4.6,1
62,1,38,38,1,0.3,91,4.3,1
73,0,67,42,1.1,0.54,66,3.7,1
81,1,15,10,0.6,0.33,141,4,1
81,1,12,6,0.6,0.37,142,3.7,1
46,1,37,29,1.2,0.5,108,3.8,1
46,1,44,30,1.4,0.53,114,3.5,1
46,1,42,22,1.5,0.7,101,3.3,1
46,1,50,76,0.4,0.18,113,4.5,1
44,1,49,31,2.6,1.1,101,3.5,1
47,1,39.5,24.5,1,0.5,293,3.2,1
47,1,25,16,1.4,0.55,124,3.9,1
59,1,32,25,1.2,0.52,91,3.5,1
77,1,15,13,0.4,0.22,63,3.9,1
80,0,36,16,2.56,1.09,124,3.5,1
63,0,50,41,1.3,0.47,85,5.2,1
49,1,34.1,31.1,0.5,0.26,82,4.7,1
69,0,19,13,0.7,0.25,80,4.7,1
72,0,27,14,0.8,0.32,111,4.1,1
85,0,64,13,0.6,0.32,140,2.7,1
```

Figure 4. Example liver ARFF file.

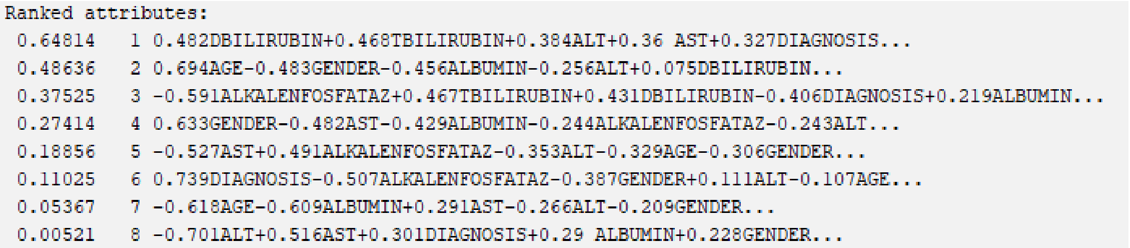
The first step in this study is to compare the performance of the ML algorithms using all features in the dataset. The obtained results for cross-validation (10-fold) and data splitting (70%-30%) techniques were summarized in Table 2 and Table 3, respectively. As seen from Table 2, RF algorithm showed higher classification outperform (with 90.75% accuracy) as compared with other algorithms as evaluated 10-fold cross validation techniques. J48, RT, K-star, REPTree and IBk algorithms, respectively, having 86%, 85.08%, 88.08%, 84.33% and 83.84% accuracy were followed RF. In addition, as seen in Table 3, the highest classification performance was also obtained for RF algorithm (89.72%) when ML performances were evaluated based data splitting method. The classification accuracies for J48, RT, K-star, REPTree and IBk algorithms were obtained 85.27%, 86.66%, 88.05%, 84.16% and 85%, respectively.

Following the analysis of the performance of ML algorithms based on all features, significant features in the dataset were identified using PCA technique. Afterwards, using these extracted features, the performances of ML algorithms were compared separately according to data data splitting and cross-validation techniques. The obtained results were given in Table 4 and Table 5 for cross-validation (10-fold) and data splitting (%70-%30) techniques, respectively.

As seen from Table 4, RF algorithm showed higher classification outperform (with 88.83% accuracy) as compared with other algorithms as evaluated 10-fold cross validation techniques. J48, RT, K-star, REPTree and IBk algorithms, respectively, having 85.75%, 84.83%, 87.91%, 84.16% and 83.91% accuracy were followed RF. In addition, as seen in Table 5, the highest classification performance was also obtained for RF algorithm (88.61%) when ML performances were evaluated based data splitting method. The classification accuracies for J48, RT, K-star, REPTree and IBk algorithms were obtained 86.11%, 85.55%, 86.38%, 82.77% and 82.77%, respectively.

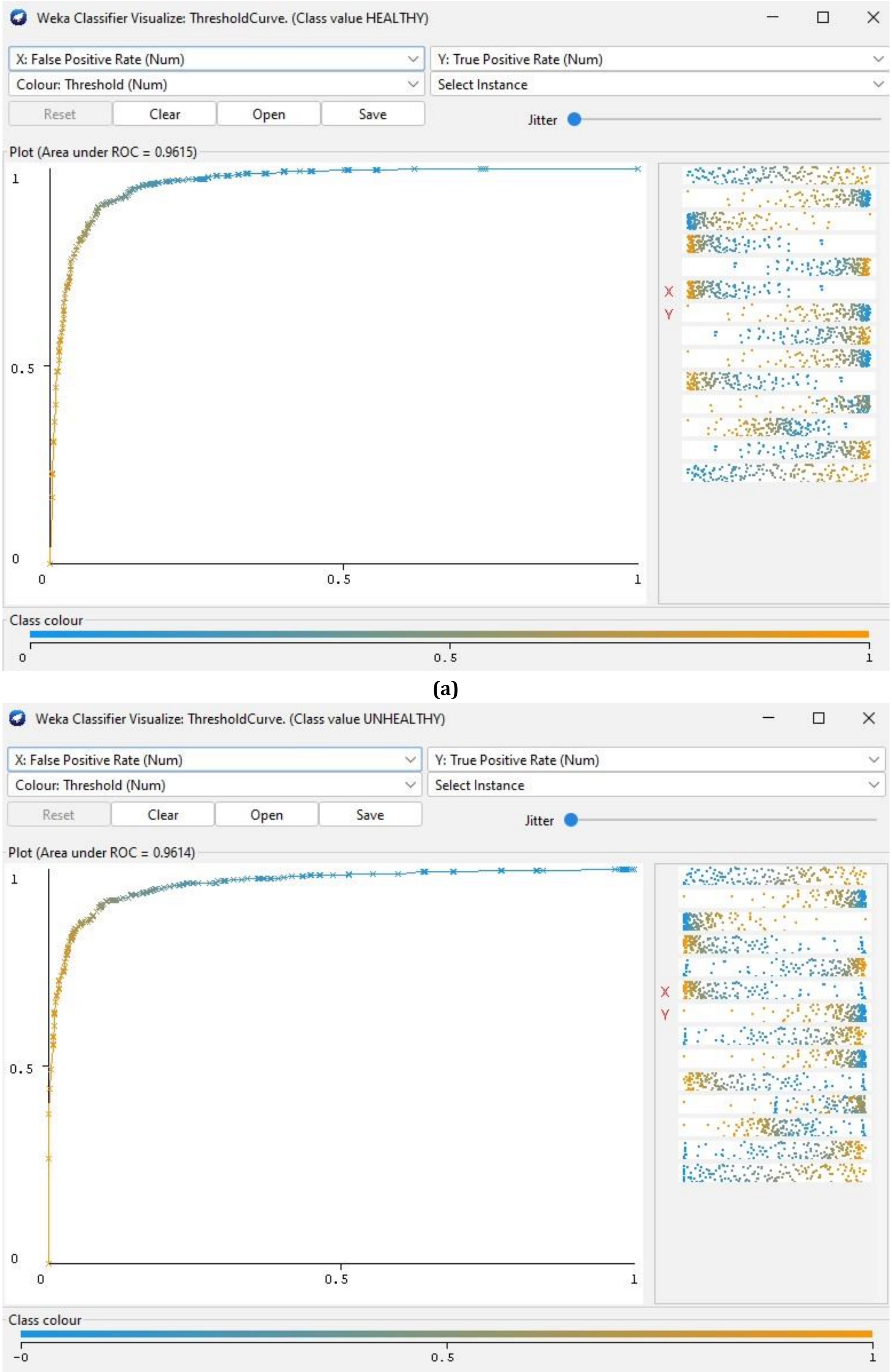
At the end of this study, ROC curves were plotted for RF algorithm having the highest classification accuracy for the first approach used all features in the dataset. ROC curves of the algorithm for 10-fold and data splitting were given in Figure 6 and Figure 7, in turn. The ROC curves of the algorithm were exhibited compatibility with the results given in Table 2 and Table 3.

Studies in the literature used ML algorithms conducted on diagnosis of liver diseases were given in Table 6. It is important to compare the results with previous studies and Table 6 summarizes this comparison. It can be clearly seen that the accuracy value obtained with proposed method is higher than other studies. The studies in the literature commonly had been performed on publicly available datasets and had not taken into consideration the specialist physicians' opinions.



```
Ranked attributes:
 0.64814   1 0.482DBILIRUBIN+0.468TBILIRUBIN+0.384ALT+0.36 AST+0.327DIAGNOSIS...
 0.48636   2 0.694AGE-0.483GENDER-0.456ALBUMIN-0.256ALT+0.075DBILIRUBIN...
 0.37525   3 -0.591ALKALENFOSFATAZ+0.467TBILIRUBIN+0.431DBILIRUBIN-0.406DIAGNOSIS+0.219ALBUMIN...
 0.27414   4 0.633GENDER-0.482AST-0.429ALBUMIN-0.244ALKALENFOSFATAZ-0.243ALT...
 0.18856   5 -0.527AST+0.491ALKALENFOSFATAZ-0.353ALT-0.329AGE-0.306GENDER...
 0.11025   6 0.739DIAGNOSIS-0.507ALKALENFOSFATAZ-0.387GENDER+0.111ALT-0.107AGE...
 0.05367   7 -0.618AGE-0.609ALBUMIN+0.291AST-0.266ALT-0.209GENDER...
 0.00521   8 -0.701ALT+0.516AST+0.301DIAGNOSIS+0.29 ALBUMIN+0.228GENDER...
```

Figure 5. Screenshot obtained by PCA analysis.



Weka Classifier Visualize: ThresholdCurve. (Class value HEALTHY)

X: False Positive Rate (Num)

Y: True Positive Rate (Num)

Colour: Threshold (Num)

Select Instance

Plot (Area under ROC = 0.9615)

(a)

Weka Classifier Visualize: ThresholdCurve. (Class value UNHEALTHY)

X: False Positive Rate (Num)

Y: True Positive Rate (Num)

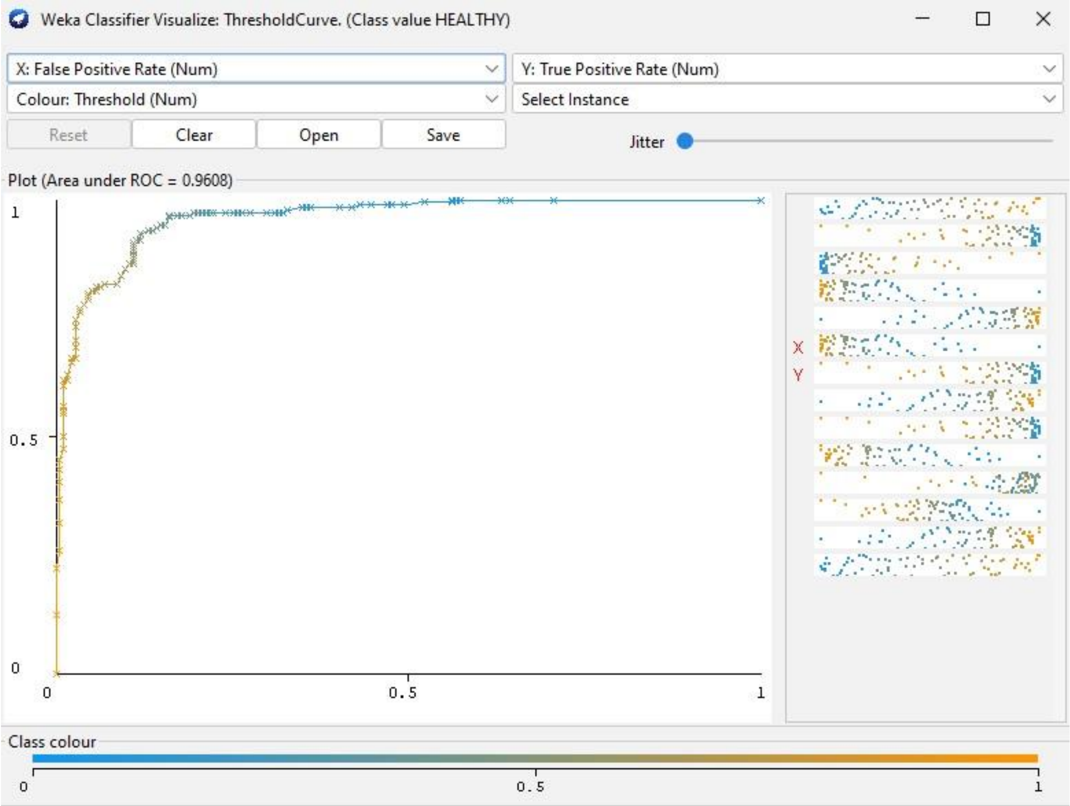
Colour: Threshold (Num)

Select Instance

Plot (Area under ROC = 0.9614)

(b)

Figure 6. ROC curves of RF algorithm for whole dataset (10-fold) a) Healthy, b) Unhealthy class.


Weka Classifier Visualize: ThresholdCurve. (Class value HEALTHY)

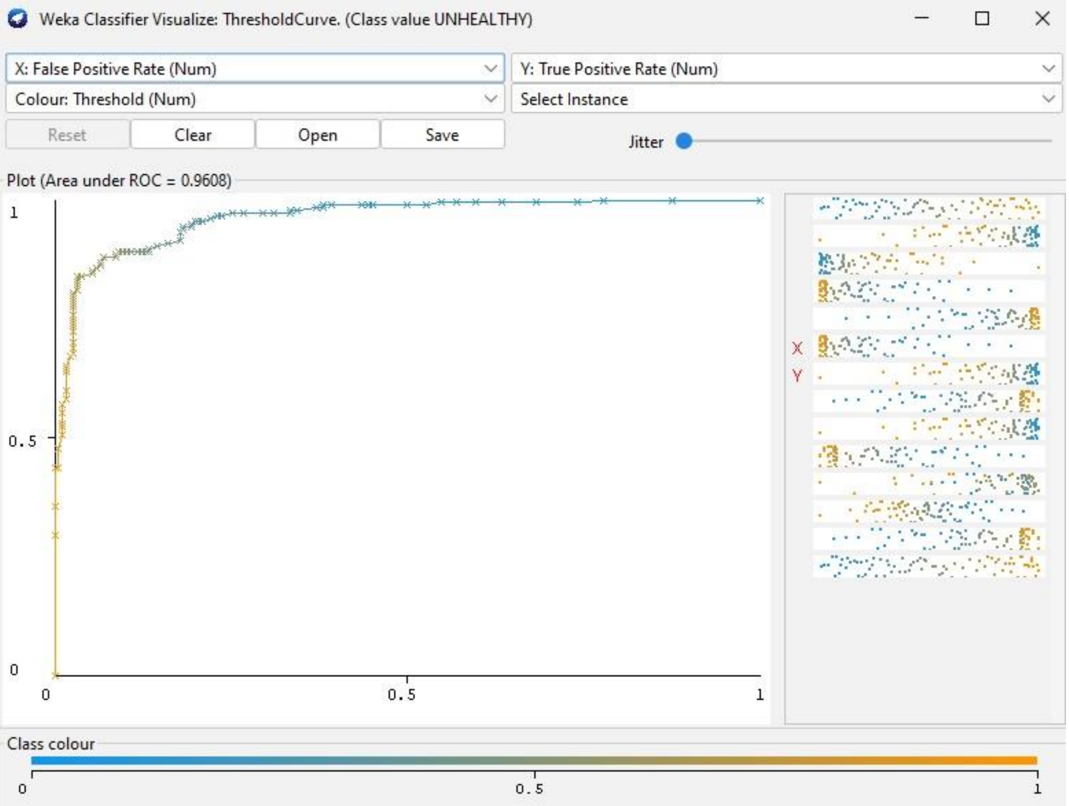
X: False Positive Rate (Num)

Y: True Positive Rate (Num)

Colour: Threshold (Num)

Plot (Area under ROC = 0.9608)

(a)


Weka Classifier Visualize: ThresholdCurve. (Class value UNHEALTHY)

X: False Positive Rate (Num)

Y: True Positive Rate (Num)

Colour: Threshold (Num)

Plot (Area under ROC = 0.9608)

(b)

Figure 7. ROC curves of RF algorithm for reduced dataset by applying PCA (70%-30%) a) Healthy, b) Unhealthy class.

Table 2. The performance of algorithms on whole dataset for 10-fold cross validation

ML Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F-Measure	ROC
IBk (k-NN)	86.66	86.80	86.70	86.70	86.50
K-star (K*)	88.08	88.10	88.10	88.10	94.40
J48 (C4.5)	86	86	86	86	86.10
Random Forest (RF)	90.75	90.80	90.80	90.70	96.10
Random Tree (RT)	85.08	85.20	85.10	85.10	84.90
REPTree	84.33	84.40	84.30	84.30	89.40

Table 3. The performance of algorithms on whole dataset for 70%-30% ratio

ML Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F-Measure	ROC
IBk (k-NN)	85	85	85	85	84.40
K-star (K*)	88.05	88.10	88.10	88.10	94.40
J48 (C4.5)	85.27	85.30	85.30	85.30	87.20
Random Forest (RF)	89.72	89.70	89.70	89.70	96.10
Random Tree (RT)	86.66	86.80	86.70	86.70	86.70
REPTree	84.16	84.20	84.20	84.20	88.30

Table 4. The performance of algorithms on reduced dataset by applying PCA for 10-fold cross validation

ML Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F-Measure	ROC
IBk (k-NN)	83.91	84	83.90	83.90	85.10
K-star (K*)	87.91	87.90	87.90	87.90	94.20
J48 (C4.5)	85.75	85.80	85.80	85.70	86
Random Forest (RF)	88.83	88.90	88.88	88.88	95.50
Random Tree (RT)	84.83	84.90	84.80	84.80	85.20
REPTree	84.16	84.20	84.20	84.20	88.50

Table 5. The performance of algorithms on reduced dataset by applying PCA for 70%-30% ratio

ML Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F-Measure	ROC
IBk (k-NN)	82.77	82.90	82.80	82.80	82.90
K-star (K*)	86.38	86.40	86.40	86.40	94
J48 (C4.5)	86.11	86.10	86.10	86.10	87
Random Forest (RF)	88.61	88.60	88.60	88.60	95.30
Random Tree (RT)	85.55	85.60	85.60	85.50	85.90
REPTree	82.77	82.80	82.80	82.80	90

5. Discussion

The aim of this study was to classify patients whose were healthy and suffer from liver fibrosis/cirrhosis (600 healthy, 600 unhealthy) using ML algorithms. The collected data set included demographic and blood values of 1200 patients. To make a binary classification of liver disease; random forest, K-star, REPTree, random tree, k-nearest neighbours and C4.5 decision tree algorithms were utilized. The performance of all ML algorithms was interpreted by applying two diverse approaches. The first approach had been realized by using all features on the dataset, whereas the second approach had been carried out using extracted features through PCA. ML performances had been assessed based on 10-fold cross-validation and data splitting (70%-30% train-test) techniques for each approach. The whole analyses were performed in WEKA. The performances of algorithms were evaluated in terms of various metrics (accuracy, precision, recall, F-measure and ROC). These metrics were preferred, because they commonly utilized in literature.

In this study, models were established based on two different approaches to diagnosing diabetes, and it was determined that the RF algorithm was the algorithm with high discrimination and the best classification criteria in general in correctly classifying the diagnosis of diabetes for both approaches.

In the light of findings, RF algorithm outperforms with high discrimination and the best classification criteria in general in correctly classifying the diagnosis of liver fibrosis/cirrhosis diseases for both approaches. Therefore, RF algorithm could be preferred to make accurate diagnosis of liver disease on the basis of this promising study results.

The studies in the literature usually aim to determine the ML algorithm with the best performance by using open access datasets. In addition, opinions of the specialist physicians have not been included in these studies. To the best of our knowledge, no comprehensive comparison of ML algorithms for the diagnosis of liver fibrosis/cirrhosis diseases has been studied on the whole dataset and PCA-reduced dataset together as in this study.

6. Conclusion

As a result, this study is expected to provide guidance in evaluating the performance of several supervised ML algorithms in liver fibrosis/cirrhosis diagnosis classification. In similar studies, if the performance of the models is to be compared, it may be recommended to evaluate them with such approaches to increase the sensitivity of the study.

In future studies, to obtain higher classification performance, it is thought to increase the dataset, to use different pre-processing techniques and to apply diverse feature selection methods gained increasingly popular recently. Performance evaluations will be investigated using different data mining methods and algorithms. The performance of the proposed model in the detection of different types of liver disease will be examined. In addition, it is planned to develop applications that can be used in the early diagnosis of fibrosis/cirrhosis and support decision-making by healthcare professionals.

Table 6. The comparison with studies-based ML algorithms on liver diseases diagnosis in the literature

Authors	Dataset	Tool	ML Algorithms	Accuracy (%)
Gulia et al. (2014)	Indian Liver Patient Dataset (ILPD)	WEKA	BayesNet, J48, MLP, Random Forest and SVM	MLP: 70.8405% SVM: 71.3551% J48: 70.669% Random Forest: 71.8696% BayesNet: 69.1252%
Muthuselvan et al. (2018)	Indian Liver Patient Dataset (ILPD)	WEKA	Naïve Bayes, k-Star, J48 and Random Tree	(After Feature Selection) Naïve Bayes: 60.6% K-star: 67.2% J48: 71.2% Random Tree: 74.2%
Keleş et al. (2020)	Indian Liver Patient Dataset (ILPD)	WEKA	J48, LMT, Decision Stump, Hoeffding Tree, REP Tree, Random Forest, Random Tree and IBk	J48: 74.40% LMT: 73.80% Decision Stump: 67.10% Hoeffding Tree: 69.10% REP Tree: 71.30% Random Forest: 81.90% Random Tree: 73.80% IBk: 81.60%
Proposed Study	Dataset collected from Zonguldak Ataturk State Hospital	WEKA	J48, IBk, K-star, Random Tree, REPTree and Random Forest (10-fold)	J48: 86% IBk: 86.66% K-star: 88.08% Random Tree: 85.08% REPTree: 84.33% Random Forest: 90.75%
Proposed Study	Dataset collected from Zonguldak Ataturk State Hospital	WEKA	J48, IBk, K-star, Random Tree, REPTree and Random Forest (10-fold after PCA Feature Selection)	J48: 85.75% IBk: 83.91% K-star: 87.91% Random Tree: 84.83% REPTree: 84.16% Random Forest: 88.83%

Author Contributions

The percentage of the author(s) contributions is presented below. All authors reviewed and approved the final version of the manuscript.

	R.U.A	Z.P.	C.K.
C	25	50	25
D	25	50	25
S	50	20	30
DCP	20	60	20
DAI	20	60	20
L	40	30	30
W	40	30	30
CR	30	30	40
SR	60		40
PM	50	25	25
FA		100	

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

This research is approved by Non-interventional Clinical Research Ethics Committee of Zonguldak Bulent Ecevit University (approval date: March 10, 2021, protocol code: 2021/05).

References

Acarlı K. 2020. Karaciğer sağlığını koruyan 10 hayati öneri. URL: https://www.memorial.com.tr/saglik-rehberi/karaciger-sagligini-koruyan-10-hayati-oneri (accessed date: August 28. 2023).

Alaybeyoğlu A, Mulayim N. 2018. Karaciğer kanseri teşhisinde destek vektör makinesi tabanli uzman sistem tasarimi. Tıp Teknolojileri Kongresi, 8-10 Kasım, Gazi Magaso, KKTC, ss: 208-210.

Alkuşak E, Gök M. 2014. Karaciğer yetmezliğinin teşhisinde makine öğrenmesi algoritmalarinin kullanimi. ISITES 2014, June 8-10, Karabük, Türkiye, pp: 703-707.

Asrani SK, Devarbhavi H, Eaton J, Kamath PS. 2019. Burden of liver diseases in the world. J Hepatol, 70(1): 151-171.

Azam MS, Rahman A, Iqbal SHS, Ahmed MT. 2020. Prediction of liver diseases by using few machine learning based approaches. Aust J Eng Innov Technol, 2(5): 85-90.

Breiman L. 2001. Random forests. Machine Learn, 45: 5-32.

Borulday MG, Yegin EG, Mahouti P, Gunes F. 2017. Diagnosing liver Diseases with decision tree algorithm. Inter J Tech Phys Problems Engin, 33: 67-70.

Bulut C, Ballı T, Yetkin EF. 2023. Filtre modelli öznitelik seçim algoritmalarının EEG tabanlı beyin bilgisayar arayüzü sistemindeki karşılaştırmalı sınıflandırma performansları. Gazi Üniv Müh Mimar Fak Derg 38(4): 2397-2408.

Del Campo JA, Gallego P, Grande L. 2018. Role of inflammatory response in liver diseases: Therapeutic strategies. World J Hepatol, 10(1): 1.

Dritsas E, Trigka M. 2023. Supervised machine learning models for liver disease risk prediction. Comput, 12(1): 19.

Duda RO, Hart PE, Stork DG. 2000. Pattern classification. Wiley, New Jersey, USA, 2nd ed., pp: 176-181.

Gaber A, Youness HA, Hamdy A, Abdelaal HM, Hassan AM. 2022. Automatic classification of fatty liver disease based on supervised learning and genetic algorithm. Applied Sci, 12(1): 521.

Gulia A, Vohra R, Rani P. 2014. Liver patient classification using intelligent techniques. Inter J Comput Sci Inform Technol 5(4): 5110-5115.

Işık K, Ulusoy SK. 2021. Metal Sektöründe üretim sürelerine etki eden faktörlerin veri madenciliği yöntemleriyle tespit edilmesi. Gazi Üniv Müh Mimar Fak Derg, 36(4): 1949-1962.

Kaya C, Erkaymaz O, Ayar O, Özer M. 2017 October. Classification of diabetic retinopathy disease from Video-Oculography (VOG): signals with feature selection based on C4. 5 decision tree. Medical Technologies National Congress (TIPTEKNO), 31 October-2 September, Trabzon, Türkiye, pp: 1-4.

Keleş A, Karslı ÖB, Keleş A. 2020. Makine öğrenme algoritmaları ile karaciğer hastalığının teşhisi. Turkish Stud Inform Technol Appl Sci, 15(1): 75-83.

Lin RH. 2009. An intelligent model for liver disease diagnosis. Artificial Intel Med, 47(1): 53-62.

Ma H, Xu CF, Shen Z, Yu CH, Li YM. 2018. Application of machine learning techniques for clinical predictive modeling: a cross-sectional study on nonalcoholic fatty liver disease in China. BioMed Res Inter, 2018: 4304376.

Mishra S, Tadesse Y, Dash A, Jena L, Ranjan P. 2021. Thyroid disorder analysis using random forest classifier. In Intelligent and Cloud Computing: Proceedings of ICICC 2019 Volume 2 Springer, Singapore, pp: 385-390.

Mukhyber SJ, Abdulah DA, Majeed AD. 2023. Classification of liver dataset using data mining algorithms. 1st International & 4th Local Conference for Pure Science (ICPS-2021), 26–27 May 2021, Diyala, Iraq, pp: 6.

Muthuselvan S, Rajapraksh S, Somasundaram K, Karthik K. 2018. Classification of liver patient dataset using machine learning algorithms. Int J Eng Technol, 7(3): 323.

Narin A, Kaya C, Pamuk Z. 2021. Automatic detection of coronavirus disease (covid-19): using x-ray images and deep convolutional neural networks. Pattern Anal Applicat, 24: 1207-1220.

Pahareeya J, Vohra R, Makhijani J, Patsariya S. 2014. Liver patient classification using intelligence techniques. Inter J Adv Res Comput Sci Software Engin, 4(2): 295-299.

Rahman AS, Shamrat FJM, Tasnim Z, Roy J, Hossain SA. 2019. A comparative study on liver disease prediction using supervised machine learning algorithms. Inter J Sci Technol Res, 8(11): 419-422.

Ramana BV, Babu MSP, Venkateswarlu NB. 2011. A critical study of selected classification algorithms for liver disease diagnosis. Inter J Database Manage Syst, 3(2): 101-114.

Schiff ER, Maddrey WC, Reddy KR. 2017. Schiff's Diseases of the Liver. John Wiley & Sons, London, UK, pp: 241.

Sevli O. 2019. Performance Comparison of Different Machine Learning Techniques in Diagnosis of Breast Cancer. Eur J Sci Technol, 16: 176-185.

Şenel FA, Saygin RR, Saygin M, Öztürk Ö. 2021. Makine Öğrenmesi Algoritmaları Kullanılarak Vücut Analizi ile Uyku Apnesi Teşhisi. Uyku Bülteni, 2(1): 6-10.

Ucar M, İncetaş MO. 2022. Classification of brain MRI using Efficientnet CNN model and feature selection method. İKSAD

Publishing House, Ankara, Türkiye, pp: 110.

Uzun R, İşler Y, Toksan M. 2018 May. Choose of wart treatment method using Naive Bayes and k-nearest neighbors classifiers. 26th Signal Processing and Communications Applications Conference (SIU), 02-05 May 2018, İzmir, Türkiye, pp: 1-4.

Uzun R, İşler Y, Toksan M. 2019. WEKA yazılım paketinin siğil tedavi yöntemlerinin başarısının tahmininde kullanımı. Düzce Üniv Bilim Teknol Derg, 7(1): 699-708.

Ünal Y, Sağlam A, Kayhan O. 2019. Improving classification performance for an imbalanced educational dataset example using SMOTE. Avrupa Bilim Teknol Derg, 2019: 485-489.

Xie G, Wang X, Liu P, Wei R, Chen W, Rajani C, Jia W. 2016. Distinctly altered gut microbiota in the progression of liver disease. Oncotarget 7(15): 19355.

Velu SR, Ravi V, Tabianan K. 2022. Data mining in predicting liver patients using classification model. Health Technol, 12(6): 1211-1235.

Yapıcı Şenyer İ. 2021. Obezitenin elektroretinografi (ERG): sinyali üzerindeki etkisi. Doktora Tezi, Zonguldak Bülent Ecevit Üniversitesi, Fen Bilimleri Enstitüsü, Zonguldak,Türkiye, ss: 163.